University of Alberta

Studies in Synthesis and Analysis of Model Predictive Controllers

by

Rohit S. Patwardhan

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

in

Process Control

Department of Chemical and Materials Engineering

Edmonton, Alberta

Fall 1999

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-46902-6

Canada

University of Alberta

Library Release Form

Name of Author: Rohit S. Patwardhan

Title of Thesis: Studies in Synthesis and Analysis of Model Predictive Controllers

Degree: Doctor of Philosophy

Year this Degree Granted: 1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Rohit S. Patwardhan
CME 536
Edmonton, AB
Canada, T6G 2G6

Date: JUNE 30, 1999.

The little I know I owe to my ignorance.

Sacha Guitry

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Studies in Synthesis and Analysis of Model Predictive Controllers** submitted by Rohit S. Patwardhan in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in *Process Control*.

S. L. Shah (Supervisor)

J. F. Forbes

T. Chen

H. Marquez

B. Huang

M. Nikolaou (External examiner)

Date: JUNE 11, 1999

To my parents

Abstract

Mode predictive control (MPC) is a paradigm that emerged in the late 1970's. Intuitive in design and computationally elegant, it has become synonymous with advanced process control in the industry for applications where interactions and constraints have to be addressed. The design phase was somewhat heuristic until Rawlings and Muske (1993) gave the tools, in the form of Lyapunov theory, for constrained stability analysis of model predictive controllers (MPC).

The purpose of this thesis is two-fold:

1. To harness the developments in numerical methods, multivariate statistics and identification techniques for improved design of model-based predictive controllers.

2. To develop tools for analysis of MBPC performance and implement them on industrial applications.

Convex optimization methods have found several applications in the area of control (Boyd *et al.*, 1994). The linear matrix inequality (LMI) framework allows one to cast several control design problems as suitable convex problems that can be solved efficiently. In this thesis the LMI framework has been used to design MPC and apply it to real-time control of pilot scale processes.

Partial least squares (PLS) gives a robust way of handling redundancies in the input-output data and coming up with truly parsimonious models. The PLS framework was integrated with the MPC methodology to conceive a novel way of designing linear and nonlinear model predictive controllers. The PLS based MPC was tested on a pilot scale process as well as a simulated, nonlinear pH-level control process.

Control relevant identification attempts to bridge the gap between identification and control objectives. The identification objective is chosen to be commensurate with the

control performance objective. In this thesis we use the finite impulse response (FIR) form to pose and solve the control relevant identification problem for model predictive control. Some steps are taken towards applying this methodology under closed loop conditions and with constraints in place.

The present day control practitioner has to spend a significant part of his/her time on maintaining the advanced control applications. Current assessment and diagnosis methods for MPC are largely heuristic and time consuming. Tools that systematically quantify MPC performance and aid in diagnosing poor performance would be of immense value to the practicing control engineer. As part of this thesis a simple way of quantifying MPC performance has been proposed and demonstrated on an industrial application. This assessment technique is comprised of comparing the designed objective function with achieved objective function and is free from any statistical assumptions such as normality or stationarity of the operating data. Uncertainty is part of the reality that most process control engineers have to face on day-to-day basis. The effect of uncertainty on controller performance is quantified as an aid to performance diagnosis. Other factors such as constraints, nonlinearity and disturbance uncertainty are also considered.

Acknowledgements

I would like to express my gratitude towards my supervisor, Prof. Shah for his constant encouragement and guidance. His emphasis on the practical aspects led me to understand the nuances of process control and discover new problems. I am grateful to him, for giving me the freedom to define my own problems, and being patient with me during unproductive times.

I would like to thank Profs. Fisher, Shah, Chen, Forbes, Guay, Huang, Wood and Rink for teaching me the fundamentals of control and the tools for doing research in my area. I feel fortunate to have learnt the subject from such fine teachers.

I would also like to express my gratitude towards Drs. Biao Huang and S. Lakshminarayanan (Laksh). I learnt the basics of performance assessment from Biao and our stimulating discussions inevitably resulted in new ideas. Laksh introduced me to the area of multivariate statistics and encouraged me to use these tools for MPC design. He was an ideal office mate and I will never forget our time at Mizushima, Japan.

I had the pleasure of interacting with Mr. Genichi Emoto and Dr. Kent Qi of Mitsubishi Chemical Co., Japan and Shell Canada, Fort Sasketchwan respectively. The industrial exposure, which included a two month stay at the Development and Engineering Research Center, Mitsubishi Chemical Co., Japan, was truly rewarding and led me to understand and appreciate the needs and problems of applied process control.

The process control group has been an exciting place to be, a place where one cannot help learning. Ravindra, Pranob, Lanre, Biao, Laksh, Dan Yang, Aseema, Anand, Amy, Lisa, Dan, Misha, Kamrun, Dongguang, Arun, Bhushan, Sachin, Ramesh and Haitao have all been friends, teachers or just sounding boards for many a stupid idea. It was my pleasure to share the office with Bhushan for the past few months. Always willing to give up the computer, I will cherish the discussions with him on control, religion, astronomy .....

My parents always encouraged me to seek out my own goals and for that I will be eternally grateful to them.

I will always remember the good times with the many many friends that I made in

Edmonton - Manav, Deepika, Janice, Parag, Yatin, Rashmi, Nehal, Anurag, Geeta, Kevin, Gagan, Sonia, Shaista.

Last but not the least I would like to place on record my gratitude towards the inhabitants of 3A-9010, 112 Street for tolerating me and letting me have all that food! Raj, Nikhil, Varada, Jin, Lucio, Jai, Eshetu and Kay - you have been wonderful roommates.

Contents

1 Introduction 1
- 1.1 Scope of this Thesis . . . 2
- 1.2 Research Objectives . . . 2
- 1.3 About the Thesis . . . 4
  - 1.3.1 About the Notation . . . 5

2 A review of model predictive control 7
- 2.1 Introduction . . . 7
- 2.2 The Control Objective . . . 9
  - 2.2.1 Constraints . . . 10
- 2.3 Predictor . . . 11
- 2.4 Unconstrained Control Law . . . 13
  - 2.4.1 State Space Form . . . 18
- 2.5 Constrained Case . . . 20
- 2.6 Nominal Design . . . 25
  - 2.6.1 Stability . . . 25
  - 2.6.2 Performance . . . 30
- 2.7 Robust MPC . . . 31
  - 2.7.1 Modeling the uncertainty . . . 31

3 Model Predictive Control via Linear Matrix Inequalities 35
- 3.1 Introduction . . . 35
- 3.2 Semidefinite Programming . . . 36
- 3.3 LMIs in Process Control . . . 37
- 3.4 A SDP Formulation of model Predictive Control . . . 39
  - 3.4.1 The unconstrained case . . . 39
  - 3.4.2 Extension to the constrained case . . . 40
- 3.5 Choice of weighting matrices via LMIs . . . 41
- 3.6 Real-time Control using LMIs . . . 42
  - 3.6.1 SISO Process . . . 42
  - 3.6.2 MIMO Process . . . 43
- 3.7 Conclusions . . . 47

4 A Partial Least Squares Framework for Model Predictive Control 49
- 4.1 Introduction to Modeling and Control via Partial Least Squares . . . 49
  - 4.1.1 PLS based Dynamic Models . . . 50
  - 4.1.2 Controller Synthesis in the PLS Framework . . . 53

4.2 Interactor Estimation and Multivariate Time Delay Compensation via PLS models . . . 54
4.2.1 Numerical Example . . . 60
4.2.2 Discussion . . . 62
4.3 Model Predictive Control in Latent Spaces . . . 64
4.3.1 PLS based constrained MPC of Linear Systems . . . 65
4.4 Constrained Nonlinear Model Predictive Control using PLS based Hammerstein and Wiener Models . . . 73
4.4.1 Nonlinear Modeling using PLS based Hammerstein and Wiener models 75
4.4.2 Constrained NLMPC via PLS based Hammerstein and Wiener Models 76
4.4.3 Discussion . . . 80
4.4.4 Case Study: pH-Level Control of an Acid-Base Neutralization process 82
4.5 Conclusions . . . 92

5 Techniques for assessment of model predictive controllers 93
5.1 Introduction . . . 93
5.2 Conventional Assessment Techniques . . . 95
5.2.1 Minimum Variance Benchmarking . . . 95
5.3 Performance Assessment of Model Predictive Controllers . . . 102
5.3.1 Properties . . . 105
5.4 Simulation examples . . . 116
5.4.1 Mixing Process . . . 116
5.4.2 Effect of reduced order modeling . . . 118
5.5 Application to QDMC assessment . . . 121
5.6 Conclusions . . . 128

6 Issues in Performance Diagnostics of Model-based Controllers 129
6.1 Introduction . . . 129
6.2 Preliminaries . . . 131
6.3 Fundamental Limitations on Performance . . . 133
6.3.1 Effects of discretization . . . 133
6.3.2 Limitations due to non-minimum phase zeros . . . 134
6.3.3 Effect of hard constraints . . . 137
6.4 Effect of uncertainty in the plant model . . . 139
6.4.1 Deterministic Inputs . . . 143
6.4.2 Mixed Inputs . . . 144
6.4.3 Uncertainty in the time delay . . . 144
6.4.4 Extension to the multivariate case . . . 145
6.4.5 Uncertainty in the noise model . . . 147
6.5 Impact of nonlinearity on performance . . . 147
6.5.1 Effect of mismatch on model predictive controllers . . . 149
6.6 Simulation examples . . . 149
6.7 Conclusions . . . 157

7 Performance Analysis of Model Predictive Controllers: Industrial Applications 159
7.1 Introduction . . . 159
7.2 Practical Methods for Performance Analysis of DMC . . . 163

7.2.1 Assessment techniques 163
7.2.2 Diagnostics 166
7.3 Case Study 1: Propylene Splitter DMC 169
7.3.1 Assessment Results 169
7.3.2 Diagnostics 174
7.4 Case Study 2: De-methanizer DMC3 182
7.4.1 Assessment Results 182
7.4.2 Diagnostics 186
7.5 Conclusions 191

8 Towards Iterative Control and Identification for MPC **193**
8.1 Introduction 193
8.2 Control Relevant Identification for MPC 196
8.3 Control Relevant Identification in the FIR form 200
8.3.1 Open loop case 201
8.3.2 MIMO extensions 203
8.3.3 Closed loop issues 204
8.4 Conclusions 210

9 Conclusions and Recommendations **212**
9.1 Concluding remarks 212
9.2 Future research directions 213

Bibliography **214**

A A method for closed loop estimation of uncertainty **222**

List of Figures

2.1 *The Internal Model Control structure* 16
2.2 *The impulse response of T* 17
2.3 *Controller pole location (o) for different control weightings (* corresponds to the plant zero)* 19
2.4 *Graphical illustration of the MPC design procedure for the unconstrained and the constrained cases.* 22
2.5 *Simulation of a distillation column under unconstrained DMC for increasing control weighting* ($p = 6, m = 1, \Gamma = I$) 25
2.6 *Simulation of a distillation column under rate plus amplitude constrained DMC* ($p = 6, m = 1, \Gamma = I, -50 \leq \Delta u(k) \leq 50, 0 \leq u(k) \leq 100$) 26
2.7 *The closed loop response to a setpoint change.* 26
2.8 *The shaded region shows the changing nature of the amplitude constraints at sampling instants (a) k=3, (b) k=4, (c) k=5 and (d) k=6.* 27

3.1 *Schematic of the light bulb experiment.* 43
3.2 *Performance of constrained LMI based MPC for the temperature control of a light bulb process.* 44
3.3 *Temperature control of a light bulb process: Effect of constraints.* 44
3.4 *Tracking performance of the constrained MPC for the temperature and level control of a mixing process* 45
3.5 *Controller performance in the regulatory mode.* 46

4.1 *Partial Least Squares based Dynamic Model* 51
4.2 *A schematic diagram showing the PLS modeling procedure.* 52
4.3 *Controller synthesis in the latent space framework* 53
4.4 *The delay in the (a) SISO case and (b) the multivariable equivalent* 55
4.5 *PLS based represenation of dynamic systems* 57
4.6 *The SISO Smith Predictor structure* 59
4.7 *The interactor factorization based multivarable dead time compensation scheme* 60
4.8 *Multivariable delay compensation in the latent space* 60
4.9 *PLS based Smith predictor for the case of no model plant mismatch (- with SP, - - without SP)* 63
4.10 *PLS based Smith predictor for the case of a delay mismatch (- with SP, - - without SP)* 64
4.11 *Linear MPC based on Dynamic PLS models* 68
4.12 *Constraint regions in the original and latent space.* 68
4.13 *Transformations of the constraint region from the original space to latent input space - necessary and sufficient conditions.* 69

4.14 *Effect of themove suppression factor on the tracking performance of the PLS based MPC (solid line) and conventional MPC (dashdotted line).* 70
4.15 *Comparison of the PLS based MPC (solid line) with PID control (dash dotted line).* 71
4.16 *Experimental evaluation of the PLS based MPC scheme on a pilot scale stirred tank heater.* 72
4.17 *Wiener Models in Latent Spaces* 76
4.18 *Hammerstein Modeling in Latent Spaces* 76
4.19 *Control strategy for PLS based Hammerstein models* 81
4.20 *Control Strategy for PLS based Wiener Models* 81
4.21 *Open-loop simulation: Level and pH responses to step changes in acid flow rate (± 1 ml/s)* 83
4.22 *Cross-Validation: Comparison of the Hammerstein model(- -) with a linear model (..) and the actual data (–)* 85
4.23 *Cross-validation fit for the Wiener model* 86
4.24 *Comparison of (1) Linear MPC (- -), (2) NLMPC-Hammerstein (–) and (3) NLMPC-Wiener (-.) for stepoint changes: pH=9, level=14.* 87
4.25 *Tracking performance of NLMPC based on (1) Hammerstein models (..) and (2) Wiener models (–) for stepoint changes: pH=4, level=18.* 88
4.26 *Response of (1) Linear MPC (-.), (2) NLMPC-Hammerstein (..) and (3) NLMPC-Wiener (–) to a change in buffer flow (0.6 to 0.2).* 89
4.27 *Performance of Wiener based NLMPC for a series of setpoint changes.* . . . 90
4.28 *Rate and Amplitude Constraints in the original space (top graph) and the latent space (bottom graph). 'x' - denotes the constraint region and 'o' denotes actual data.* 91

5.1 *A conventional control system with unmeasured disturbances and measurement noise* 95
5.2 *The closed loop impulse response coefficients for the two controller gains.* . 97
5.3 *A graphical illustration of the performance index calculation.* 97
5.4 *Illustration of the desired settling time based benchmark.* 98
5.5 *The optimal performance curve obtained through LQG benchmarking* 102
5.6 *MPC performance assessment using the LQG benchmark* 117
5.7 *Comparison of the achieved performance with the LQG objective function.* 118
5.8 *The input moves for the constrained controller during the regulatory run* . . 119
5.9 *Regulatory performance of a model predictive controller.* 120
5.10 *Comparison of design (–) and achieved performance objectives (- -) for $\lambda = 0$ (a) and $\lambda = 0.5$ (c) and the corresponding servo responses achieved by MPC - (b) and (d)* 121
5.11 *A schematic overview of the process* 122
5.12 *The step response models between the active inputs and outputs.* 124
5.13 *Performance Measure based on comparison of the design and achieved objective functions.* 125
5.14 *The FFs time series plot.* 126
5.15 *The mean values of the design (solid line) and the achieved objective functions (dotted line) shown along with their components - J_y, J_u, J_p and J_a.* 127

5.16 *Performance indices based on comparison of the different components in the objective function.* 127

6.1 *The conventional feedback system* 131
6.2 *The internal model control stucture* 134
6.3 *Illustration of the effect of modelling uncertainty on achieved performance* . 141
6.4 *The performance and robustness properties of the control system with a unit delay mismatch.* 150
6.5 *Comparison of achieved and designed closed loop step responses for the case where (a) the delay is underestimated and (b) the delay is overestimated.* . 152
6.6 *The 1-norm of the controller transfer function Q, against increasing controller gains.* 152
6.7 *The controller response (u) for K=0.2 (solid line), K=0.3 (dash-dotted line) and K=0.4 (dotted line)* 153
6.8 *The phase and magnitude characteristics of the (multiplicative) uncertainty in time delay.* 153
6.9 *Nominal closed loop step response of a H_2 optimal controller* 155
6.10 *The effect of a structural mismatch on the performance and stability of a MPC controller. (a) The design and achieved performance for Q_1 and (b) the robustness properties of Q_1 and Q_2.* 156
6.11 *Effect of uncertainty in the interactor on the optimal cost function for LQG.* 156
6.12 *Effect of control weighting on robustness margin.* 157

7.1 *Schematic of the DMC control structure. The controller in the conventional sense is shown as the dynamic control module with the DMI model in the feedback path. The LP part at the top sets the targets for the controlled and the manipulated variables.* 161
7.2 *A representative diagram showing the various steps in performance analysis.* 162
7.3 *The effect of gain mismatch on the linear programming step in DMC.* . . . 167
7.4 *Scatter plot showing a characteristic template for valve hysteresis.* 167
7.5 *The control heirarchy in a chemical plant.* 168
7.6 *A schematic diagram of the distillation column showing the relevant controlled and manipulated variables.* 170
7.7 *The Production requirements from July 1 to Oct. 28. The data sets corresponded to the highlighted regions.* 170
7.8 *The input-output data for the normal load conditions* 171
7.9 *The multivariable minimum variance performance index and the historical benchmark, for the (i) raw and (ii) weighted data.* 172
7.10 *The achieved objective and the contributing factors for the high load data.* . 173
7.11 *(a) The MVC performance indices for different data sets (1-Normal, 2-High, 3-Low A, 4-Low-B) and (b) the historical performance measures.* 173
7.12 *The input-output data for the low load (B) data set. The reflux flow rate is at the lower operating limit.* 174
7.13 *The internal model control structure for DMC* 175
7.14 *Prediction errors for the DMI model: low load.* 176
7.15 *Prediction errors for the DMI model: high load.* 176

7.16 *The top graph shows the output errors of y2 vs. y1 and the bottom graph shows the corresponding prediction errors. The circles (o) denote the data from the low load conditions while the crosses (+) denotes high load data. Every sixth point from the high load data set is numbered.* 177
7.17 *The prediction errors for low load data set A.* 178
7.18 *The prediction errors for low load data set B.* 178
7.19 *Measures for prediction errors based on the minimum variance benchmark.* . 179
7.20 *The finite impulse response models between the top impurity and the reflux flow rate. These models were estimated for the low and high load conditions using univariate correlation analysis (after prewhitening) and are compared with DMI model.* 180
7.21 *Output error (OE) models between the top impurity and the reflux flow rate with and without pre-whitening. The effect of feedback is clearly seen when no pre-whitening is done.* 181
7.22 *Multivariate performance measures for the demethanizer DMC based on the settling time index.* 183
7.23 *Individual performance measures for output 2, along with the output control error trends.* 184
7.24 *The univariate performance index for output 4.* 185
7.25 *The pre-fractionator bottoms impurity shown along with the setpoint.* 185
7.26 *Performance assessment for the PID loops.* 186
7.27 *Closed loop pole locations for the temperature loop.* 187
7.28 *The closed loop step responses for a period of time spanning over 11 days.* . 188
7.29 *The LP targets along with the outputs for the column pressure and the reboiler level.* 188
7.30 *Interpretation of the DMC structure including the LP step.* 189
7.31 *The feedback representaion of DMC LP step.* 190

8.1 *Magnitude response of $L^2(q^{-1})$ for different weightings* 199
8.2 *Comparison of control relevant FIR identification with conventional least squares method. True impule response (o), control relevant model (*) and conventional least squares estimate ($\triangle$)* 204
8.3 *The feedback structure* 205
8.4 *The magnitude response of the pre-whitening filter (dashed line) and the sensitivity function (solid line).* 206
8.5 *Estimation of the step response coefficients using correlation analysis under closed loop conditions with pre-whitening (dash-dotted) and without pre-whitening (dotted) shown with the true step response (solid).* 206
8.6 *The effect of pre-whitening on the OE structure: OE with pre-whitening (dotted line), without pre-whitening (dashdotted line) and the true response (solid line).* 207
8.7 *The frequency domain fits for the OE models with and without pre-whitening. OE with pre-whitening (dotted line), without pre-whitening (dashdotted line) and the true response (solid line).* 207
8.8 *Effect of constraints on closed loop identification: direct identification with constraints (dashed line), without constraints (dashdotted line). The solid line gives the true step response.* 209

8.9 *Effect of constraint activation on closed loop identification.* 209
8.10 *Effect of valve nonlinearity on closed loop identification schemes.* 210

A.1 *The IMC structure* 223
A.2 *Comparison of the estimated mismatch (dashed line) with the actual mismatch (solid line).* 223

List of Tables

3.1 *Computational Effort for simulation casestudy (m=2)* 48
3.2 *Computational Effort for the MIMO case* 48

5.1 *Effect of constraints on MPC performance* 118
5.2 *Effect of prediction horizon on regulatory performance* 124
5.3 *Tracking and regulatory performance* 124
5.4 *Disturbance rejection properties of MPC* 124
5.5 *Contribution of various terms to the design and achieved objective functions* 126

7.1 *Performance measures and the prediction measures for the propylene splitter DMC application.* 183
7.2 *Summary of performance diagnosis results for the propylene splitter DMC application.* 184

Chapter 1

Introduction

In the past two decades process control practice has progressed from simple PID controllers to advanced model based controllers capable of handling interactions and constraints. On the other hand, control research has evolved from the state-space techniques such as LQG to robust synthesis techniques such as H_∞ and control relevant identification methods which attempt a synergistic treatment of the identification and control problem. Researchers have also attempted to harness these advances in linear control theory to nonlinear processes. As a control researcher, one is often concerned with the *synthesis* of a controller for a particular process. In other words, one would like to know what kind of inputs have to be designed so that the system behaves in a pre-specified manner. Thus controller synthesis is really concerned with the design of the controller. If we turn the problem around and fix the inputs to the system, we are now interested in studying the behavior of the system. This is known as *analysis*. *Analysis* attempts to investigate the structure and behavior of the system. To a control practitioner, the analysis of an existing system provides valuable insight into the behavior of the control system.

Model predictive controllers (MPC) belong to a class of model-based controllers which compute future control actions by minimizing a performance objective function over a finite prediction horizon. In a receding horizon fashion, they implement only the first of the calculated control moves. These controllers usually have a step/impulse response model to predict the future plant response. This family of controllers is truly multivariate in nature and has the ability to deal with constraints on the inputs, slew rates, *etc.* It is for the above reasons that MPC has been widely accepted by the process industry. Variations of MPC such as dynamic matrix control (DMC) and quadratic dynamic matrix control (QDMC) have become the norm in industry for processes where interactions are of foremost importance and constraints have to be taken into account.

The field of model predictive control emerged in the late 1970s and has held centre stage in the area of process control mainly due to its industrial successes. Researchers have looked at developing tuning guidelines, a variety of formulations (state-space, transfer function, internal model control, *etc.*), extension to nonlinear processes, and in the 90's the focus has been on robust design of MPC. Methods such as the min-max approach and the l_1 optimal theory are now available for robust design of multivariate constrained model predictive controllers.

1.1 Scope of this Thesis

In this work we study a wide class of techniques for design and analysis of linear and nonlinear model predictive controllers. In more general terms MPC belongs to the class of discrete controllers and can be single or multi-rate. In an industrial setting the challenge is two-fold:

- Design a high performance model predictive controller for a complex process based on, mostly, an empirical model estimated from input-output data.
- Monitor, maintain and if possible improve the performance of existing MPC applications.

The design phase involves (i) identifying an appropriate model from input-output data, (ii) choice of tuning parameters to optimize performance and provide stability, (iii) on-line implementation - which involves solving a potentially large optimization problem at every sampling instant.

The monitoring and maintenance phase can split into three distinct tasks: (i) obtaining performance metrics from operating data and design information, (ii) in case the performance metric indicates less than satisfactory performance, establishing the underlying causes (iii) determining the corrective actions to be taken for improving performance. The last step is tantamount to re-design of the existing application and can even entail re-identifying the process model. In a practical setting, the improvement of performance will be achieved through some iteration between the design and analysis steps.

1.2 Research Objectives

More specifically the research objectives of this project included:

Development of novel ways of synthesizing model predictive controllers and their evaluation on pilot scale processes

1. *Use of convex optimization methods for MPC design.*

   Convex optimization methods are very reliable and efficient tools for control system design. Interior point methods are polynomial time algorithms for solving convex programs with guaranteed convergence to the global minima. We propose the use of the linear matrix inequality (LMI) framework to solve the optimization step in MPC via interior point methods.

2. *Use of partial least squares (PLS) based dynamic models for MPC.*

   The PLS framework is particularly useful in dealing with colinear data. Constrained model predictive control of linear as well as nonlinear processes, based on dynamic PLS models is proposed and demonstrated on simulation as well as pilot scale processes.

3. *Development of control relevant identification for model predictive control.*

   Until recently the identification step was considered as independent step with little or no relation to the control objective. In order to overcome the discrepancies between the identification and control objectives one must resort to control relevant identification. The FIR form is naturally suited for MPC design. The control relevant identification problem is posed and solved in the FIR form. It is one of the aims of this thesis to study the properties of this identification algorithm in the context of closed loop issues.

Development of tools for performance analysis of model predictive and model based controllers in general

1. *Development of performance assessment techniques for MPC.*

   The performance assessment of constrained model predictive controllers presents a challenging research problem. Constrained MPC is essentially a linear time-varying control law due to the changing nature of the active constraints. Current performance assessment techniques compare performance against a benchmark which is typically unrealizable. This is not very useful in diagnosing poor controller performance. A more practical approach is to compare achieved performance with the designed performance. The MPC design method offers a natural choice of such a design criterion- the design objective function.

2. *Development of methods for diagnosing poor performance in model based controllers.*

   A poor performance index can be the outcome of one or more of the following: (1) Poor controller design (2) Persistent disturbances (3) Limits on achievable performance with existing process control structure (4) Design constraints (5) Changes in plant dynamics. How does one decide what is the contribution of each of these terms towards poor performance? We propose to quantify the contributions of these factors towards degradation in performance.

3. *Industrial application of the proposed MPC performance analysis tools.*

   Maintenance of model-based predictive controllers is an important part of a control engineer's job in the process industry. It is one of the aims of this thesis to demonstrate the suitability of the assessment and diagnosis techniques on industrial model predictive controllers such as DMC or QDMC.

1.3 About the Thesis

As outlined in the earlier section this thesis deals with the development of synthesis and analysis tools for model-based predictive controllers. To make the reader familiar with the MPC concepts and terminology, Chapter 2 begins with an overview of model predictive control. Key stability results are reviewed and the recent developments in robust MPC design are surveyed. This chapter endeavours to explain the MPC philosophy in a tutorial manner and it forms the requisite background for all of the remaining seven chapters.

Chapters 3, 4 and 8 deal with MPC design. In Chapter 3, the LMI framework for MPC design is presented. This enables the solution of the optimization step in MPC via the powerful interior point methods. The LMI based MPC is demonstrated in real time on pilot scale processes. A new tuning method based on optimizing the condition of an appropriate matrix is proposed.

Chapter 4 discusses the application of PLS based dynamic models for the design of linear and nonlinear model predictive controllers. PLS based dynamic models are reviewed and the role of constraints is highlighted. The proposed PLS based MPC is applied on a nonlinear pH-level control problem.

Chapter 8 deals with control relevant identification for MPC. The area of control relevant identification is reviewed in the context of MPC. The FIR form is utilized for estimating the MPC relevant model. Some closed loop issues such as pre-filtering, constraints and nonlinearities are also addressed.

Chapters 5, 6 and 7 focus on the development of analysis techniques for maintaining model-based predictive controllers and their application to industrial applications. In Chapter 5 a new method for assessing MPC is proposed. This method returns a performance metric based on comparison the design objective function with the achieved objective function. Some useful properties of this performance metric are proved and its usefulness is further demonstrated on a QDMC application.

Chapter 6 is step towards developing diagnostic tools for model based controllers in general. Towards this end the effect of uncertainty on performance of a control system is established. The effect of uncertainty on performance measures is quantified. This chapter also discusses the impact of constraints, nonlinearity and disturbance uncertainty on controller performance.

Chapter 7 presents an industrial case study consisting of two DMC applications at the Mizushima plant of the Mitsubishi Chemical Co., Japan. Several practical measures of performance are developed and applied in a case study manner. The role of the lower control layers in analyzing DMC performance is emphasized. Some structural aspects of DMC are highlighted, with particular attention to stability and performance issues of the control system.

The significant contributions of this thesis are outlined in Chapter 9 and future research directions are indicated.

The material is presented with a tutorial flavour so as to familiarize the reader with the diverse fields such as linear matrix inequalities, partial least squares, performance assessment, iterative control and identification, covered in this thesis. Since this thesis is written in a paper format, some degree of overlap could not be avoided. For example, the figure explaining the IMC structure occurs more than once.

1.3.1 About the Notation

The notation used in this thesis, with the exception of Chapter 4, is consistent with the MPC literature and the control literature in general. For Chapter 4, since it deals with a multivariate statistical method - PLS, the notation is introduced at the start. The Z-transform is used throughout the thesis to represent the discrete system in the frequency domain. For the sake of completeness, we give the definition of the Z-transform here

$$y(z) = \sum_{k=0}^{\infty} y(k) z^{-k} \tag{1.1}$$

where z is a complex variable. The forward shift, backward shift and the difference operators are defined as

$$\begin{aligned} qy(k) &= y(k+1) \\ q^{-1}y(k) &= y(k-1) \\ \Delta y(k) &= (1-q^{-1})y(k) = y(k) - y(k-1) \end{aligned} \tag{1.2}$$

respectively. The symbol $\succeq$ is used to denote the positive semidefiniteness of a matrix, $P \in R^{n \times n}$

$$\begin{aligned} P &\succeq 0 \Leftrightarrow \\ x^T P x &\geq 0, \forall x \neq 0, x \in R^n. \end{aligned}$$

Conventional vector norms used in this thesis are the Euclidean norm or the 2-norm:

$$||x||_2^2 = x^T x,$$

the 1-norm:

$$||x||_1 = \sum_{k=1}^{n} |x_k|$$

and the ∞-norm:

$$||x||_\infty = \max_k |x_k|.$$

Chapter 2

A review of model predictive control

2.1 Introduction

The term MPC represents a family of model based predictive controllers. The MPC family algorithms are designed on the basis of a multistep optimization objective. In general, several controller moves in the future are computed but only the first control action is implemented, hence these controllers are also referred to as receding horizon controllers. The earlier versions of MPC are - the identification and command algorithm (IDCOM) proposed by Richalet *et al.* (1978) and the dynamic matrix control (DMC) algorithm due to Cutler and Ramaker (1980). Other well known variations of MPC include: model algorithmic control (MAC) by Rouhani and Mehra (1982) , multivariable optimal constrained control algorithm (MOCCA) from Sripada and Fisher (1986) and generalized predictive control (GPC) due to Clarke *et al.* (1987) .

The underlying philosophy of MPC type control algorithms differs from conventional PID controllers in several respects:

- An explicit model of the process is used within the control algorithm to determine the control actions at every step based on the minimization of a cost function.
- This model is often in the finite impulse response (FIR) form which can be obtained with relative ease from a large chemical plant. The FIR model has the ability to capture complex dynamics.
- The formulation is truly multivariate in nature and can handle interactions in a consistent manner.

- This family of controllers has the ability to deal with hard constraints on the inputs and outputs in an optimal way. This represents a significant step in terms of practical implementation.

- The computational complexity of the optimization step is restricted to a linear or quadratic program in the worst case. Thus these algorithms can be easily implemented on-line.

- At each sampling instant several control actions are calculated, only the first control move is implemented. These controllers are thus known as *receding horizon* controllers.

The basic ideas behind model predictive control predate the methods proposed by Richalet *et al.* (1978) and the DMC algorithm due to Cutler and Ramaker (1980). Garcia *et al.* (1989) and Bannerjee (1996) give a historical background for the MPC concept which originated in the early 1960's. Simplicity of design combined with its ability to tackle realities such as constraints and interactions has helped MPC achieve its current popularity with the process industry. The industrial successes of these algorithms spurred the growth of MPC as a research area in academia. Researchers have analyzed the MPC algorithm in great detail. MPC has been reformulated and analyzed in the internal model control framework (Garcia and Morari, 1982), the linear quadratic gaussian framework (Bitmead *et al.*, 1990) and also in the stochastic, input-output setting via generalized predictive control (Clarke *et al.*, 1987).

In this chapter, we give an overview of the key ideas involved in classical model predictive control with a tutorial flavor. Section 2.2 illustrates the control objective in its commonly used forms and introduces the MPC terminology. Constraints are also highlighted with respect to their type and geometry. Standard model forms for predicting the process response are compared in Section 2.3. The derivation of a closed loop expression for the unconstrained predictive control law is given in Section 2.4, in the transfer function and the state space forms. An illustrative example is used to demonstrate the effect of the tuning parameters on the controller stability. The constrained solution for a quadratic objective function and linear constraints is outlined in Section 2.5. Section 2.6 summarizes methods for nominal design and analysis of MPC. The nominal performance issues are discussed in light of the tracking and disturbance rejection properties of predictive controllers. Incorporating modeling uncertainty into the design of a MPC is crucial to its success. The recent developments in the design of robust MPC are covered in Section 2.7.

2.2 The Control Objective

The underlying philosophy of model predictive control (MPC) consists of minimization of a performance objective function with respect to future input moves, over a finite time horizon. The standard objective function in model based predictive controllers consists of the sum of (i) weighted norm of the control errors over a prediction horizon, p and (ii) a weighted norm of the control moves over a control horizon, m,

$$J_{QMPC} = \sum_{i=1}^{p} ||r(k+i) - \hat{y}(k+i/k)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||\Delta u(k+i-1)||^2_{\Lambda_i} + \sum_{i=1}^{m} ||u(k+i-1)||^2_{R_i} \tag{2.1}$$

here $r \in R^{n_y}$ is the reference signal, $\hat{y} \in R^{n_y}$ is the predicted output, $u \in R^{n_u}$ is the control input, $\Gamma_i \succeq 0, \Lambda_i \succ 0, R_i \succ 0$ are the respective weighting matrices. $||.||$ is used to denote the Euclidean norm (2-norm) of a vector. In equation 2.1 the 2-norm is weighted with appropriate matrices. Another commonly used objective function is based on minimizing the 1-norm of the relevant signals,

$$J_{LMPC} = \sum_{i=1}^{p} |r(k+i) - \hat{y}(k+i/k)|_{\Gamma_i} + \sum_{i=1}^{m} |\Delta u(k+i-1)|_{\Lambda_i} + \sum_{i=1}^{m} |u(k+i-1)|_{R_i}. \tag{2.2}$$

In MPC terminology, p is called the prediction horizon; m is known as the control horizon; Γ_i is the output weighting matrix; Λ_i is the input move suppression factor; R_i is the input weighting matrix. In recent literature the infinite horizon MPC has received much attention ($p \to \infty$). If the model is linear, the convexity of the objective function is guaranteed, since Λ_i is positive definite. Other modifications of the objective function include formulation in terms of the states instead of the input-output form and inclusion of the intersample behaviour. These objective functions are shown here for the regulatory case,

$$J_{SS} = \sum_{i=1}^{p} ||x(k+i/k)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||\Delta u(k+i-1)||^2_{\Lambda_i} + \sum_{i=1}^{m} ||u(k+i-1)||^2_{R_i} \tag{2.3}$$

$$J_C = \sum_{i=1}^{p} \int_{t=(k+i-1)h}^{(k+i)h} ||x(t)||^2_{\Gamma_i} dt + \sum_{i=1}^{m} ||\Delta u(k+i-1)||^2_{\Lambda_i} + \sum_{i=1}^{m} ||u(k+i-1)||^2_{R_i}.$$

The control objective is to compute a sequence of control moves that minimizes one of the above objective functions. In reality the inputs are limited to a region of allowable moves, $\Omega_k \subset R^m$. Thus in practice, the following constrained optimization problem is solved at every sampling instant,

$$\min_{u_k \in \Omega_k} J(u)$$

where $u_k = [u(k)...u(k+m-1)]^T$. This is a finite dimensional optimization problem. However, after computing the optimal move sequence u_k^*, only the first of these optimal moves is implemented on the process. This is known as receding horizon control. The remaining moves are discarded and a new optimization problem is solved at the next sampling instant,

$$u(k) = u_k^*(1) = u^*(k).$$

2.2.1 Constraints

A real plant has to work within certain physical limitations, e.g. a valve which can handle only a particular range of flow rates, and market forces which result in rigid quality requirements on the process outputs. Usually a real process involves rate and amplitude constraints on the input, and may also require output constraints to be considered,

$$\begin{aligned} \Delta u_{\min} &\leq \Delta u(k) \leq \Delta u_{\max} : \text{ Rate Constraints} \\ u_{\min} &\leq u(k) \leq u_{\max} : \text{ Amplitude Constraints} \\ y_{\min} &\leq \hat{y}(k) \leq y_{\max} : \text{ Quality Constraints.} \end{aligned}$$

Often times the manipulated variables may have an undesirable impact on some key process variables. This effect can be minimized through the incorporation of constraints on auxiliary variables,

$$f_{\min} \leq \hat{f}(k) \leq f_{\max} : \text{ Quality Constraints on auxillary variables.}$$

The constraints are not limited to the linear set. For example the 2-norm of the input/output vectors can be constrained. A contraction constraint on the states can also be included to ensure nominal stability,

$$\begin{aligned} ||u||_2^2 &\leq \sigma_u^2 \\ ||\hat{y}||_2^2 &\leq \sigma_y^2 \\ ||x(k+p/k)||_2 &\leq \alpha ||x(k+1)||_2. \end{aligned}$$

Equality constraints may include terminal constraints on the state or the output,

$$\begin{aligned} \hat{y}(k+p/k) &= y_{des} \\ x(k+p/k) &= 0. \end{aligned}$$

We define a region $\Omega_k \subseteq R^{m.n_u}$ such that if $u_k \in \Omega_k$ all the specified constraints on the inputs and outputs are satisfied. Depending upon the model form used, the constraints can

define a convex or a non-convex feasible region, Ω_k. We define a region $\Omega_k \subseteq R^{m.n_u}$ such that if $u_k \in \Omega_k$ all the specified constraints on the inputs and outputs are satisfied. For a linear model, the constraints of the above form lead to time-varying, convex Ω_k.

Bemporad and Morari (1998) incorporate logical constraints into a predictive control framework. The logical constraints may include physical rules, if-then-else rules, on-off switches. The inclusion of logical constraints leads to presence of integer or binary (0/1) variables in the constraint set. They propose the use of mixed integer quadratic programming (MIQP) to solve the resulting mixed logical dynamical system thus arising.

2.3 Predictor

The key component of any model predictive control scheme is a predictor which is capable of accurately predicting the process response over a prediction horizon, based on the current and past measurements. Different model forms can be used to predict the process behaviour. The commonly used model forms include: (i) the step response or the finite impulse response model, (ii) the state space form and (iii) the transfer function form. The general model forms for each of these are tabulated below.

$$\begin{aligned}
\hat{y}(k) &= \sum_{j=1}^{N} s(j)\Delta u(k-j) + s(N)u(k-N-1) + e(k) : \text{ Step Response} \\
&= \sum_{j=1}^{N} h(j)u(k-j) + e(k) : \text{ FIR} \\
\hat{y}(k) &= C(Ax(k-1) + Bu(k-1)) + e(k) : \text{ State Space} \\
A(z)\hat{y}(k) &= B(z)u(k) + \frac{C(z)}{\Delta(z)}e(k) : \text{ Transfer function}
\end{aligned}$$

where $e(k)$ is a white noise process. The FIR model is equivalent to using an output error (OE) model in the transfer function domain. The transfer function form given here is a discrete autoregressive integrated moving average with exogenous input (ARIMAX) model. The i-step ahead predictors for the respective model forms are compared below,

$$\begin{aligned}
\hat{y}_d(k+i/k) &= \sum_{j=1}^{N} s(j)\Delta u(k+i-j) + s(N)u(k+i-N-1) \qquad (2.4) \\
\hat{y}(k+i/k) &= CA^i\hat{x}(k) + CA^iBu(k) + \ldots + CABu(k+i-1) \\
\hat{y}(k+i/k) &= G_i(q)\Delta u(k+i-1) + \frac{\tilde{G}_i(q)}{C(q)}\Delta u(k-1) + \frac{F_i(q)}{C(q)}y(k)
\end{aligned}$$

where $G_i(q), \bar{G}_i(q)$ and $F_i(q)$ satisfy the following Diophantine identities (Bannerjee, 1996):

$$\begin{aligned} \frac{C(q)}{\Delta A(q)} &= E_i(q) + q^{-i}\frac{F_i(q)}{\Delta A(q)} \\ \frac{E_i(q)B(q)}{C(q)} &= G_i(q) + q^{-i}\frac{\bar{G}_i(q)}{C(q)}. \end{aligned}$$

The effect of unmeasured disturbances is incorporated to some extent in the FIR form through a feedback term,

$$\begin{aligned} \hat{y}(k+i/k) &= \hat{y}_d(k+i/k) + d(k+i/k) \\ d(k+i/k) &= \hat{y}(k) - y(k). \end{aligned}$$

The full state measurements are rarely available. Hence a state estimator or observer accompanies a state space based predictor. The state space model (Lee *et al.*, 1994; Li *et al.*, 1989) will include a disturbance model indirectly through a Kalman filter or any other observer used for estimating the state,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + B_1 w(k) \\ y(k) &= Cx(k) + v(k) \\ \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + L[y(k) - \hat{y}(k)]. \end{aligned} \tag{2.5}$$

The state space formulation for the predictor has some desirable properties such as the presence of an optimal filter for incorporating the effect of unmeasured disturbances and the ability to handle the multivariate case. The optimal i-step ahead predictor for the transfer function model involves solution of certain Diophantine equations and also incorporates an explicit model for unmeasured disturbances. The transfer function approach does not extend easily to the multivariate and non-square case. Both the state space and the transfer models can predict the response of unstable/integrating processes whereas the FIR form cannot. One common way of accommodating the integrating outputs in the FIR form is to formulate the prediction in terms of the differenced output which is the same as factoring out the integrator,

$$\begin{aligned} y(k) &= \frac{B}{A\Delta}u(k) \Leftrightarrow \\ \Delta y(k) &= \frac{B}{A}u(k) = \sum_{j=1}^{N} h(j)u(k-j). \end{aligned} \tag{2.6}$$

The differenced output is included in the control objective instead of the original variable. The control of level is an example of the integrating output and is often handled by controlling the changes in the level.

In a novel approach Bannerjee (1996) used Laguerre polynomials to represent the deterministic form of the input-output model,

$$\hat{y}(k) = \sum_{i=1}^{N} g_i \Psi_i(q) u(k) \tag{2.7}$$

where $\Psi_i(q)$ are the orthonormal basis functions of Laguerre type:

$$\Psi_i(q) = \frac{\sqrt{1-a^2}}{q-a} \left(\frac{1-aq}{q-a} \right)^{i-1}. \tag{2.8}$$

Laguerre function models with a real pole a, ($|a| < 1$) are used to represent the dynamics of an overdamped process. Since Markov functions belong to a special class of orthonormal functions, the FIR model is known as a Markov model.

2.4 Unconstrained Control Law

The formulation of the model predictive controller for the MIMO case is discussed here before expressing the unconstrained control law as a linear feedback controller. The model assuming (i) linearity and (ii) an open loop stable process is:

$$\hat{y}_k = S\Delta u_k + f_k + d_k \tag{2.9a}$$

where the individual terms are given by:

$$S = \begin{bmatrix} s_1 & 0 & & \cdots & 0 \\ s_2 & s_1 & 0 & \cdots & 0 \\ s_3 & s_2 & s_1 & 0 & 0 \\ \vdots & \vdots & & \ddots & \\ s_p & s_{p-1} & & \cdots & s_{p-m+1} \end{bmatrix}_{n_y.p \times n_u.m}$$

$$\begin{aligned} \Delta u_k &= [\Delta u_1(k) \ldots \Delta u_{n_u}(k) \ldots \Delta u_1(k+m-1) \ldots \Delta u_{n_u}(k+m-1)]^T \\ \hat{y}_k &= [\hat{y}_1(k+1/k) \ldots \hat{y}_{n_y}(k+1/k) \ldots \hat{y}_1(k+p/k) \ldots \hat{y}_{n_y}(k+p/k)]^T \\ f_k &= [f_1(k+1) \ldots f_{n_y}(k+1) \ldots f_1(k+p) \ldots f_{n_y}(k+p)]^T \\ d_k &= d(k)\,[I \ldots . I]^T. \end{aligned}$$

Each s_i is a matrix of dimension $n_y \times n_u$ consisting of the step response coefficients of the process model. The second term on the right hand side denotes the free response of the system *i.e.* the response if there were no further changes in the inputs. This free response

is expressed in terms of the past (N) inputs and can be written as

$$f_k = H_1 \Delta u_o + H_2 u_o \tag{2.10}$$

$$H_1 = \begin{bmatrix} s_2 & s_3 & s_4 & \dots & & s_N \\ s_3 & s_4 & \dots & & s_N & 0 \\ s_4 & s_5 & \dots & & & \\ \vdots & & & & & \vdots \\ s_{N_2} & \dots & & s_N & & 0 \\ s_{N_2+1} & \dots & s_N & 0 & \dots & 0 \end{bmatrix}, H_2 = s_{ss} I, \tag{2.11}$$

$$\Delta u_o = [\Delta u(k-1) ... \Delta u(k-N+1)]^T, \tag{2.12}$$

$$u_o = [u(k-N) ... u(k-N+p-1)]^T. \tag{2.13}$$

The third term in equation 2.9a is the disturbance estimate $d(k)$, which is found by subtracting the model output from the actual output:

$$d(k) = y(k) - \hat{y}(k). \tag{2.14}$$

The disturbance is assumed to be constant over the prediction horizon. The objective function for MPC type controllers describes the future deviations from the setpoint p moves into the future and the differenced input moves, m moves into the future, both squared and weighted,

$$J = (r_k - \hat{y}_k)^T \Gamma (r_k - \hat{y}) + \Delta u_k^T \Lambda \Delta u_k$$

where $r_k = r(k)[I....I]^T$ is the setpoint assumed to be constant over the prediction horizon. In the above formulation the setpoint is assumed to be constant over the prediction horizon, in general a trajectory can be computed for the future setpoint. $\Gamma = diag(\Gamma_1, ..., \Gamma_p), \Lambda = diag(\Lambda_1, ..., \Lambda_m)$ are the weighting matrices defined over the entire prediction and control horizons respectively. Γ, Λ are positive semidefinite matrices usually diagonal in nature with the individual weights indicating the relative significance of the different inputs/outputs over the relevant horizons. The sequence of steps leading to the minimization of the objective function with respect to the future input moves Δu_k is outlined below,

$$\begin{aligned} J &= (r_k - S\Delta u_k - f_k - d_k)^T \Gamma (r_k - S\Delta u_k - f_k - d_k) + \Delta u_k^T \Lambda \Delta u_k \\ &= r_k^T \Gamma r_k + \Delta u_k^T (S^T \Gamma S + \Lambda) \Delta u_k - 2(r_k - f_k - d_k)^T \Gamma S \Delta u_k. \end{aligned}$$

Differentiating the objective function with respect to the future control moves we get:

$$\begin{aligned} \frac{\partial J}{\partial \Delta u} &= 2(S^T \Gamma S + \Lambda) \Delta u_k - 2S^T \Gamma (r_k - f_k - d_k) = 0 \Leftrightarrow \\ \Delta u_k^* &= (S^T \Gamma S + \Lambda)^{-1} S^T \Gamma (r_k - f_k - d_k). \end{aligned} \tag{2.15}$$

In receding horizon fashion only the first of the m moves, is implemented. Thus

$$\begin{aligned}\Delta u(k) &= [\ I\ \ 0\ \ \ldots\ \ 0\ ](S^T\Gamma S+\Lambda)^{-1}S^T\Gamma(r_k-f_k-d_k) \\ &= K(r_k-f_k-d_k)\end{aligned} \tag{2.16}$$

The conditioning of (S^TS) is critical to the implementation of model predictive controller. A poorly conditioned S^TS means that the controller will make large control moves to compensate for the effect of small disturbances or mismatch. In other words the controller will be very aggressive. A well conditioned S^TS implies that the controller will be more circumspect and make moves in proportion to the modelling errors or disturbances. Qi (1997), Qi and Badgewell (1997) discuss ways to improve the conditioning of S^TS based on matrix factorization techniques such as" singular value decomposition.
Substituting equations 2.10 and 2.14 into equation 2.16 we get,

$$\begin{aligned}\Delta u(k) &= K\left[r-H_1\Delta u_o-H_2u_o-[I....I]^T(y(k)-\hat{y}(k))\right] \\ &= K[I....I]^T(r(k)-y(k)+\hat{y}(k))-K(H_1\Delta u_o+H_2u_o) \\ &= K_1[e(k)+\hat{y}(k)]-K(H_1\Delta u_o+H_2u_o).\end{aligned}$$

Taking Z-transform on both sides we get

$$\begin{aligned}\Delta U(z^{-1}) &= K_1[E(z^{-1})+\hat{Y}(z^{-1})]-K\{H_1[\ z^{-1}\ \ z^{-2}\ \ \ldots\ \ z^{-N+1}\ ]^T\Delta U(z^{-1}) \\ &\quad +H_2[\ z^{-N}\ \ z^{-N+1}\ \ \ldots\ \ z^{-N+p-1}\ ]^TU(z^{-1})\} \\ &= K_1[E(z^{-1})+\hat{Y}(z^{-1})]-KH_1(z^{-1})\Delta U(z^{-1})-KH_2(z^{-1})U(z^{-1}).\end{aligned}$$

Letting $H(z)=(1-z^{-1})(I+KH_1(z^{-1}))+KH_2(z^{-1})$ we have the linear control as

$$H(z^{-1})U(z^{-1})=K_1[R(z^{-1})-Y(z^{-1})+\hat{Y}(z^{-1})]. \tag{2.17a}$$

This control law can be explicitly stated as

$$U(z^{-1}) = Q(z^{-1})[R(z^{-1})-Y(z^{-1})+\hat{Y}(z^{-1})] \tag{2.18}$$

$$Q(z^{-1}) = H^{-1}(z^{-1})K_1. \tag{2.19}$$

As a result of penalizing the differenced input term in the control objective the predictive controller can track step changes in the setpoint asymptotically. The controller transfer function derived here is in the internal model control framework (Garcia and Morari, 1982). Q is a high order ($=N$, number of step response coefficients) controller. The conventional

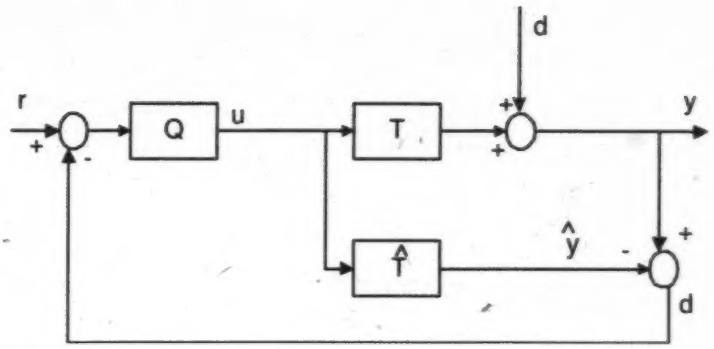


Figure 2.1: *The Internal Model Control structure*

feedback controller which acts on the control errors is related to the IMC controller in figure 8.11 by

$$C = Q(1 - \hat{T}Q)^{-1}. \tag{2.20}$$

If the future setpoints and/or the future disturbances are not assumed to be constant, a setpoint filter and/or a filter in the feedback path have to be incorporated. Garcia and Morari (1982) analyze the effect of the following exponential filter for the feedback term:

$$\hat{d}(k+i/k) = \alpha\hat{d}(k+i-1/k) + (1-\alpha)\hat{d}(k/k). \tag{2.21}$$

The following results from Garica and Morari (1982) give some important properties of the unconstrained predictive controller for the SISO, open loop stable case.

Theorem 2.1. *For $\Gamma \neq 0, \Lambda = 0$,selecting $p = m \leq N$ yields the model inverse controller $Q = T^{-1}$*

Theorem 2.2. *Assume that the system has a discrete monotonic step response and that $\Gamma = I, \Lambda = 0, p = N$. There exists a $m \ll p$ such that the controller Q is stable.*

Theorem 2.3. *There exists a $\Lambda^* > 0$ such that for all $\Lambda \geq \Lambda^*$, the predictive controller, Q, is stable for all $m \geq 1, p \geq 1$,and $\Gamma \geq 0$.*

Example 2.1. *Consider the following discrete time, SISO process:*

$$T = \frac{(1 - 0.2z^{-1})}{(1 - 0.05z^{-1})}.$$

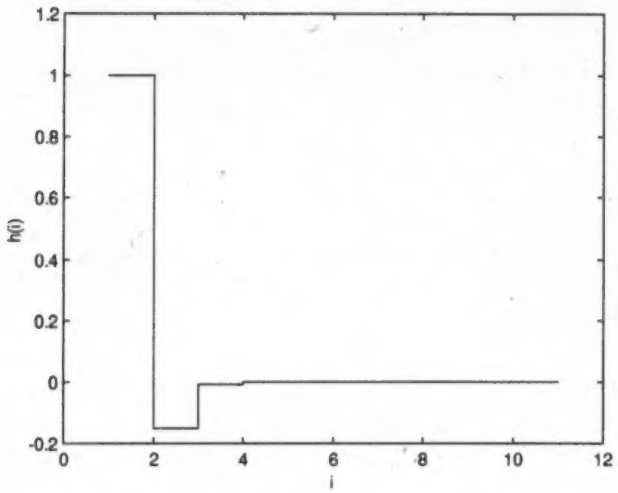


Figure 2.2: *The impulse response of T*

The impulse response of T is shown in figure 2.2.A truncated FIR form of T is given by

$$T = 1 - 0.15z^{-1} - 0.0075z^{-2} - 0.0004z^{-3}$$

For $M = P = 4$, the dynamic matrix is given by:

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.85 & 1 & 0 & 0 \\ 0.8425 & 0.85 & 1 & 0 \\ 0.8421 & 0.8425 & 0.85 & 1 \end{bmatrix}$$

and the associated matrices for the prediction of free response are given by ($N = 5$):

$$H_1 = \begin{bmatrix} 0.85 & 0.8425 & 0.8421 & 0.8421 \\ 0.8425 & 0.8421 & 0.8421 & 0 \\ 0.8421 & 0.8421 & 0 & 0 \\ 0.8421 & 0 & 0 & 0 \end{bmatrix}, H_2 = \begin{bmatrix} 0.8421 & & & \\ & 0.8421 & & \\ & & 0.8421 & \\ & & & 0.8421 \end{bmatrix}.$$

The controller gain matrix,K ($\Lambda = 0, \Gamma = I$), is given by:

$$\begin{aligned} K &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} (G^T G)^{-1} G^T \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} G^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

and

$$KH_1 \begin{bmatrix} z^{-1} & z^{-2} & z^{-3} & z^{-4} \end{bmatrix}^T = 0.85z^{-1} + 0.8425z^{-2} + 0.8421z^{-3} + 0.8421z^{-4}$$

$$KH_2 \begin{bmatrix} z^{-5} & z^{-4} & z^{-3} & z^{-2} \end{bmatrix}^T = 0.8421z^{-5}$$

$$K_1 = K \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T = 1$$

$$H(z^{-1}) = (1 - z^{-1})\{1 + KH_1(z^{-1})\} + KH_2(z^{-1})$$

$$= 1 - 0.15z^{-1} - 0.0075z^{-2} - 0.0004z^{-3}$$

Note that H is the finite impulse response (FIR) representation of T. Thus the controller is given by

$$Q_1 = \frac{1}{H(z^{-1})} = T^{-1}.$$

This is a stable controller since T is a minimum phase transfer function. Thus we have an illustration of theorem 1. Let us derive the transfer function for a non-zero control weighting, $\Lambda = I$. Repeating the above calculations yields the following controller:

$$Q_2 = \frac{K_1}{H(z^{-1})} = \frac{0.6324}{1 - 0.4643z^{-1} - 0.0031z^{-2}}.$$

The controller is no longer equal to the plant inverse, but notice that the controller gain is equal to the inverse of the plant gain, $Q_2(1) = 1.1871 = 1/0.8421 = 1/T(1)$. Also the effective controller is just second order. For $\Lambda = 0$ the controller was third order. Increasing the control weighting further to 10 gives the following controller,

$$Q_3 = \frac{K_1}{H(z^{-1})} = \frac{0.2291}{1 - 0.8091z^{-1} - 0.0006z^{-2}}.$$

Figure 2.3 shows a plot of the dominant controller poles for different control weightings. Notice that the controller pole moves towards the unit circle with increasing control weighting i.e. the controller becomes sluggish with increasing Λ. Since the process is open loop stable, a stable Q guarantees closed loop stability. It is also possible to express the predictive control law in the form of an integral controller using equation. 8.11.

2.4.1 State Space Form

Here we start with a state space from of the process model,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Gw(k) \\ y(k) &= Cx(k) + v(k) \end{aligned} \tag{2.22}$$

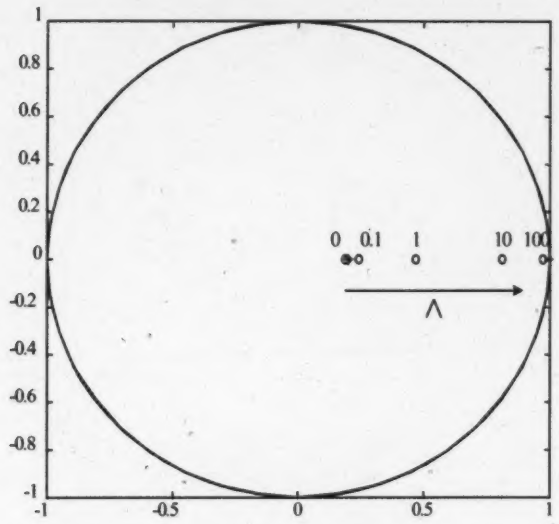


Figure 2.3: *Controller pole location (o) for different control weightings (* corresponds to the plant zero)*

where $w(k), v(k)$ are white noise processes. The predictions can be expressed in terms of the future inputs and current state variables as

$$\begin{aligned}
\hat{y}(k+1/k) &= CA\hat{x}(k) + CBu(k) && (2.23)\\
\hat{y}(k+2/k) &= CA^2\hat{x}(k) + CABu(k) + CBu(k+1)\\
&\vdots\\
\hat{y}(k+m/k) &= CA^m\hat{x}(k) + CA^mBu(k) + \ldots + CBu(k+m-1)\\
&\vdots\\
\hat{y}(k+p/k) &= CA^p\hat{x}(k) + CA^pBu(k) + \ldots + CA^{p-m-2}Bu(k+m-2)\\
&\quad + \sum_{i=1}^{p-m} CA^{p-m-i}Bu(k+m-1). && (2.24)
\end{aligned}$$

In matrix form the above predictions become

$$\begin{aligned}
\hat{y}_k &= \begin{bmatrix} CA \\ CA^2 \\ \\ CA^p \end{bmatrix} \hat{x}(k) + S\Delta u_k && (2.25)\\
&= F\hat{x}(k) + S\Delta u_k && (2.26)
\end{aligned}$$

where S is once again the dynamic matrix made up of the step response coefficients. The objective function remains unchanged except for the *rephrasing* of the free response in

terms of the current state instead of the past $(N+1)$ inputs. The disturbance estimate now becomes,

$$d(k) = y_m(k) - y(k) = C[x(k) - \hat{x}(k)] \tag{2.27}$$

where $\hat{x}(k)$ is the state estimate obtained through a Kalman predictor/filter. Thus the objective function now becomes,

$$\begin{aligned} J &= \{r_k - F\hat{x}(k) - S\Delta u_k - d\}^T \Gamma \{r_k - F\hat{x}(k) - S\Delta u_k - d_k\} + \Delta u_k^T \Lambda \Delta u_k \quad (2.28) \\ &= r_k^T \Gamma r_k + \Delta u_k^T (S^T \Gamma S + \Lambda) \Delta u_k - 2(r_k - F\hat{x}(k) - d_k)^T \Gamma S \Delta u_k. \end{aligned}$$

Once again minimization gives us the same result, albeit in a different form,

$$\begin{aligned} \therefore \Delta u(k) &= K(r_k - F\hat{x}(k) - d_k) \quad (2.29) \\ &= K[I...I]^T (r(k) - y(k)) - K\{F - C[I \ldots I]^T\}\hat{x}(k) \\ &= K_1 e(k) - K_2 \hat{x}(k). \end{aligned}$$

For the nominal case the feedback term will correspond to the noise term.

2.5 Constrained Case

A real plant has to work within certain physical limitations, e.g. a valve which can handle only a particular range of flow rates, and market forces which result in rigid quality requirements on the process outputs. A model predictive controller is well suited to handle such constraints in an elegant manner. The constraints on the inputs, slew rates and outputs can be grouped together into a single matrix inequality. Usually a real process involves rate and amplitude constraints on the input, and may also require output constraints to be considered,

$$\begin{aligned} \Delta u_{\min} &\leq \Delta u(k+i) \leq \Delta u_{\max}, i = 0, ..., m-1 \quad (2.30) \\ u_{\min} &\leq u(k+i) \leq u_{\max}, i = 0, ..., m-1 \\ y_{\min} &\leq \hat{y}(k+i/k) \leq y_{\max}, i = 1, ..., p. \end{aligned}$$

It is important that the controller satisfy these constraints. Through some standard matrix algebra we can we define these constraints together in the form of a single inequality. If we

define the following matrices,

$$L_1 = [\Delta u_{\min} \dots \Delta u_{\min}]^T ; U_1 = [\Delta u_{\max} \dots \Delta u_{\max}]^T$$

$$L_2 = \begin{bmatrix} u_{\min} - u(k-1) \\ \vdots \\ u_{\min} - u(k-1) \end{bmatrix} ; U_2 = \begin{bmatrix} u_{\max} - u(k-1) \\ \vdots \\ u_{\max} - u(k-1) \end{bmatrix}$$

$$L_3 = [y_{\min} \dots y_{\min}]^T - f - d; U_3 = [y_{\max} \dots y_{\max}]^T - f - d$$

$$R = \begin{bmatrix} I & 0 & & 0 \\ I & I & \ddots & \\ & \ddots & \ddots & 0 \\ I & & I & I \end{bmatrix}$$

then in terms of these vectors/matrices the constraints can be rewritten as:

$$\begin{array}{l} L_1 \le \Delta u \le U_1 : \text{Rate} \\ L_2 \le R\Delta u \le U_2 : \text{Amplitude} \\ L_3 \le S\Delta u \le U_3 : \text{Output.} \end{array} \tag{2.31}$$

If we let,

$$\begin{array}{l} A = [\ -I \quad -R \quad -S \quad I \quad R \quad S\ ]^T \\ B_k = [\ L_1 \quad L_2 \quad L_3 \quad -U_1 \quad -U_2 \quad -U_3\ ]^T \end{array} \tag{2.32}$$

then the constraint set (Equation. 2.31) can be represented in terms of the following inequality,

$$A\Delta u + B_k \ge 0. \tag{2.33}$$

Thus the objective function is minimized subject to linear constraints on the future inputs. An optimization problem, *e. g.* quadratic program, is solved at every sampling instant to arrive at the input values to be implemented on the plant. *A priori* one cannot say which set of constraints maybe active at a sampling instant. In some cases only the slew rate constraints are active, in other cases both slew rate and input constraints are active. Though the constrained solution is obtained by solving a set of linear equations, the changing nature of the active constraint set renders the overall control law *piecewise linear.*

The quadratic program solved at every instant is

$$\min_{\Delta u} J = \Delta u^T (S^T \Gamma S + \Lambda)\Delta u - 2(r - f - d)^T \Gamma S \Delta u \tag{2.34}$$

$$= \frac{1}{2}\Delta u^T H e \Delta u - c^T \Delta u \tag{2.35}$$

$$\text{s. t. } A\Delta u + B_k \ge 0. \tag{2.36}$$

The free response vector f depends on the past $(N+1)$ inputs. Thus the objective function has a fixed Hessian, He, but a time-varying c. The constraint region is also time varying. For $m = 2$, the objective function contours will be ellipsoidal with a changing origin at each sampling instant. Let (A^*, B_k^*) denote the active constraint set and λ be the corresponding Lagrangen multipliers. Therefore the constrained solution is obtained by solving the following set of linear equations

$$\begin{bmatrix} He & A^{*T} \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \Delta u \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ -B_k^* \end{bmatrix}. \tag{2.37}$$

The overall control law can expressed as

$$\Delta u = f(u(k-1), ..., u(k-N), y(k), r(k)). \tag{2.38}$$

The constrained control law is linear time-varying and can be shown to be Lipschitz continuous. The geometry of the constrained optimum is shown in figure 2.4 for $m = 2$. The receding horizon implementation is clearly suboptimal. Mutha (1990) attempted to obtain the constrained solution, analytically, for small horizons. Since the constrained optimization involves a potentially exponential search for the active constraint set, the numerical efficiency of MPC can be improved by first checking if the unconstrained optimum indeed satisfies the constraints.

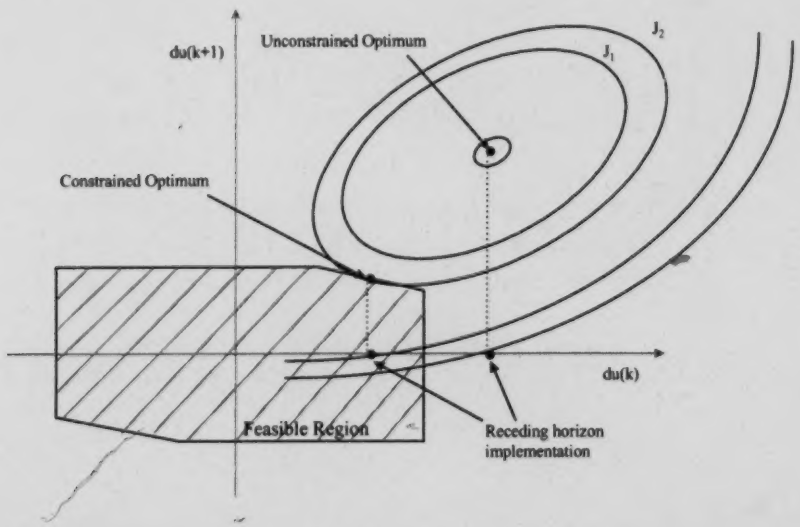


Figure 2.4: *Graphical illustration of the MPC design procedure for the unconstrained and the constrained cases.*

The output constraints cannot always be satisfied. The presence of these constraints can

lead to infeasibility. One common corrective measure is to introduce slack variables so as to relax the output constraints temporarily,

$$y_{\min} - \varepsilon \leq \hat{y}(k) \leq y_{\max} + \varepsilon. \tag{2.39}$$

It is for this reason that the output/state constraints are known as soft constraints. Rawlings and Muske (1993) give lower and upper limits on the prediction horizon necessary for satisfaction and feasibility of the state constraints. Other remedies include penalizing the constraint violations over the prediction horizon in the objective function itself,

$$\begin{aligned} J &= (r_k - \hat{y}_k)^T \Gamma (r_k - \hat{y}_k) + \Delta u_k^T \Lambda \Delta u_k \\ &\quad + (y_{\max} - \hat{y}_k)^T Q_1 (y_{\max} - \hat{y}_k) + (\hat{y}_k - y_{\min})^T Q_2 (\hat{y}_k - y_{\min}). \end{aligned} \tag{2.40}$$

The last two terms are included if the predicted outputs are violating either the lower or the upper bounds. In commercial implementation an output constraint is often left out of the formulation if it leads to infeasibility. A priority ranking is used to decide the order in which the constraints should be dropped out of the formulation in order to restore feasibility. In an application of stochastic programming Schwarm and Nikolaou (1998) solved the MPC optimization under uncertainty. The following variables were defined:

$$\begin{aligned} \varepsilon_{\max}(k+i/k) &= y_{\max} - \hat{y}(k+i/k) \\ \varepsilon_{\min}(k+i/k) &= \hat{y}(k+i/k) - y_{\max}. \end{aligned} \tag{2.41}$$

A term comprising the constraint violations weighted with the probability of constraint violations is penalized in the objective function. The constraint violations are assumed to be normally distributed. This is a limiting assumption since it holds if the modeling errors are described by a white noise process.

Example 2.2. *Consider the following continuous time model of a distillation column. The transfer function of the distillation column with the top composition (x_d) as the output and the reflux flow rate (R) as the manipulated variable:*

$$\frac{X_d(s)}{R(s)} = \frac{1.0081 e^{-0.35s}}{4.36s + 1}.$$

The sampling time was chosen as 0.35 minutes. The tuning parameters for MPC were set to $P = 6, M = 2, \lambda = 0$. The reflux rate is manipulated using a valve which has physical

limitations:

$$0 \leq u(k) \leq 100\%$$
$$-50 \leq \Delta u(k) \leq 50\%.$$

To demonstrate the effect of constraints, simulation were performed for unconstrained MPC, rate constrained MPC, amplitude constrained MPC and rate plus amplitude constrained MPC. For each of these simulations, the setpoint was a square wave between 30 and 50 units. The discretized model using a zero-order hold is given by

$$\frac{X_d(z^{-1})}{R(z^{-1})} = \frac{0.0778z^{-2}}{1.0000 - 0.9229z^{-1}}.$$

For the above tuning parameters the dynamic matrix becomes

$$S = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0.0778 & 0 \\ 0.1495 & 0.0778 \\ 0.2158 & 0.1495 \\ 0.2769 & 0.2158 \end{bmatrix}$$

At $k = 1$, the unconstrained law is given by

$$\begin{aligned} \Delta u &= (S^T S)^{-1} S^T (r_1 - f_1) \\ &= \begin{bmatrix} 119.72 & -165.51 \\ -165.51 & 242.17 \end{bmatrix} S' [\ 30 \ \ 30 \ \ 30 \ \ 30 \ \ 30 \ \ 30\ ]^T \\ &= \begin{bmatrix} 119.72 & -165.51 \\ -165.51 & 242.17 \end{bmatrix} \begin{bmatrix} 21.59 \\ 13.29 \end{bmatrix} \\ &= \begin{bmatrix} 385.79 \\ -356.03 \end{bmatrix}. \end{aligned}$$

In receding horizon fashion only the first move is implemented, $u(1) = u(0) + \Delta u(1) = 385.79$. From fig. 6.7 we can see that this is in fact the first control move implemented by the controller. The response to this input move can be found as:

$$\begin{aligned} y(3) &= 0.9229y(2) + 0.0778u(1) \\ &= 0 + .0778 * 385.79 \\ &= 30.01 \end{aligned}$$

This is deadbeat or the model inverse control. Fig.6.7 also shows the effect of the λ as a tuning parameter. As we increase λ the controller becomes more sluggish and the controller is no longer the model inverse. It has been shown by researchers that this parameter plays an important role in determining the stability of the MPC control law.

Constrained MPC: *. For the case where the unconstrained control law does not violate any constraints, the solution to the constrained case remains the same as the unconstrained one. But as we saw in the previous Section the first unconstrained control move is calculated as* $u(1) = 385.79$ *which cannot be implemented under the existing physical limitations on the valve-amplitude constraints. Nor is it possible to make such a large change in the valve opening in a single step-rate constraints. Figure 2.6 illustrate how the input profile changes as a result of the imposed constraints. For rate plus amplitude constraints the first move is* $u(1) = 50$*. Figures 2.7 and 2.8 illustrate the time varying nature of the constraints and the objective function. The unconstrained solution is indicated by '*' and the constrained solution is shown by a diamond and the receding horizon solution by 'o'.*

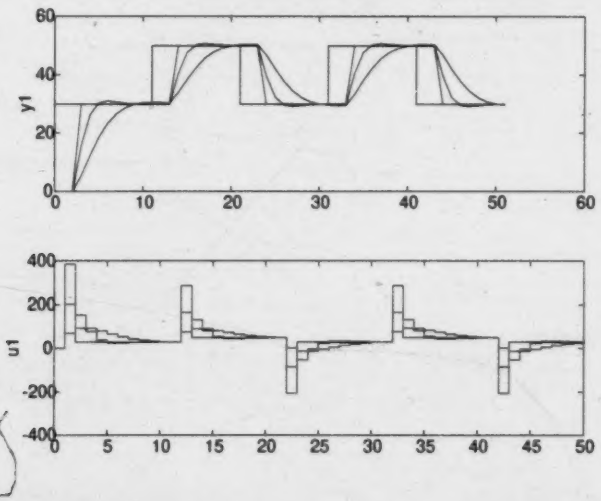


Figure 2.5: *Simulation of a distillation column under unconstrained DMC for increasing control weighting* ($p = 6, m = 1, \Gamma = I$)

2.6 Nominal Design

2.6.1 Stability

The nominal stability of MPC algorithms has been analyzed by researchers for the unconstrained and the constrained cases - Garcia and Morari (1982) , Clarke *et al.* (1987), Clarke and Mohtadi (1989), Zafiriou (1990), Zafiriou and Marchal (1991). For the unconstrained case a closed form expression can be derived (see Section 2.4) for the controller and stability analysis is straightforward. Bitmead *et al.* (1990) show the stability properties of finite

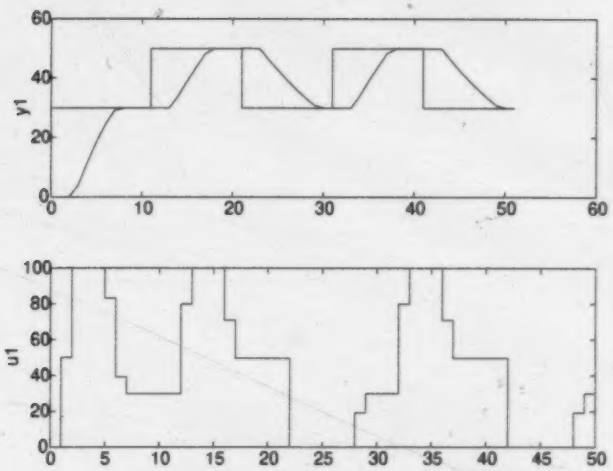


Figure 2.6: *Simulation of a distillation column under rate plus amplitude constrained DMC* ($p = 6, m = 1, \Gamma = I, -50 \leq \Delta u(k) \leq 50, 0 \leq u(k) \leq 100$)

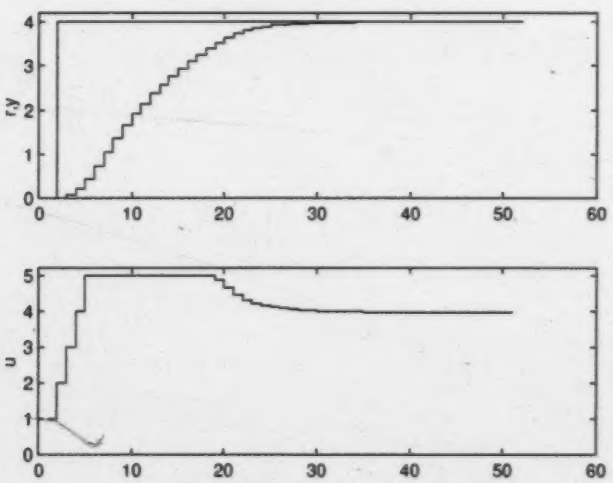


Figure 2.7: *The closed loop response to a setpoint change.*

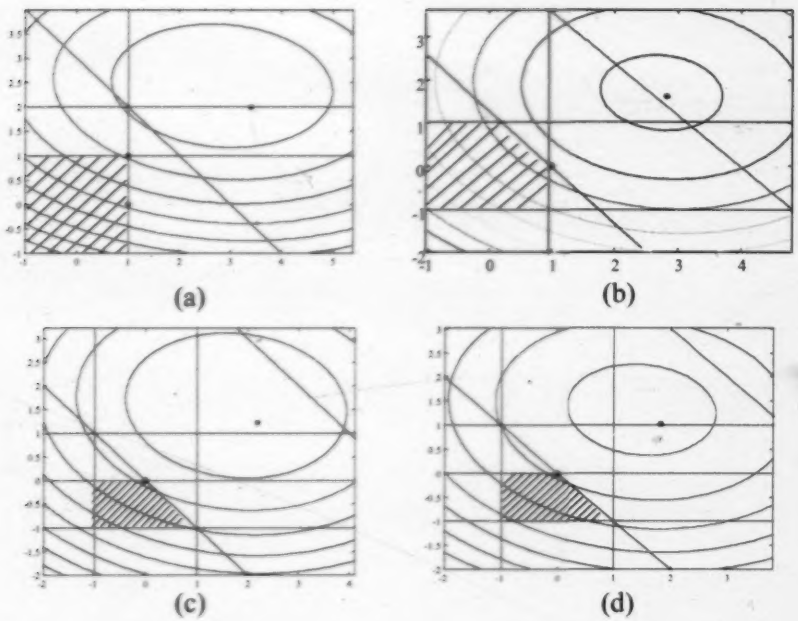


Figure 2.8: *The shaded region shows the changing nature of the amplitude constraints at sampling instants (a) k=3, (b) k=4, (c) k=5 and (d) k=6.*

horizon MPC to be poor . For the constrained case Zafiriou (1990) used the contraction mapping principle to show that the closed loop system was indeed contractive irrespective of the active constraint set. This involved checking the contraction condition for all possible combinations of the active constraints and becomes a intractable problem for large constraint sets.

A survey of the recent literature shows that the following four approaches have emerged to design stabilizing model based predictive controllers. The first and the most popular technique is that of terminal state or output weighting. The optimal input sequence has to be computed such that the states at the end of the prediction horizon are required to go the origin. Kwon and Pearson (1977) showed that this approach guarantees nominal closed loop stability,

$$\begin{aligned} & \min_{u_k} \sum_{i=1}^{P} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{M} ||\Delta u(k+i-1)||^2_{\Lambda_i} \qquad (2.42) \\ x(k+P) &= 0 \\ u_k &\in \Omega_k. \end{aligned}$$

The same effect can be achieved by putting a large weighting on the terminal outputs (Kwok and Shah, 1994). For this method to work, the system needs to be controllable and the input horizon should be larger than or equal to the state dimension. The stability is proven using Lyapunov's theorem.

For the second approach an infinite prediction horizon and a finite control horizon optimization problem is solved at every sampling instant. This method, first proposed by Rawlings and Muske (1993), is known to guarantee nominal stability of open loop stable processes for input and state constrained systems. The optimization problem is redefined as follows,

$$\min_{u_k \in \Omega_k} \sum_{i=1}^{\infty} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||u(k+i-1)||^2_{\Lambda_i}. \qquad (2.43)$$

Asymptotic stability is proven using Lyapunov's theorem. The optimal cost function, φ_k, is shown to be non-increasing. Since φ_k is bounded below by zero, φ_k reaches a constant as $k \to \infty$. Thus $x(k) \to 0$ and $u(k) \to 0$ at large times. Muske and Rawlings (1993) extended this approach to the case of unstable plants and showed that a terminal constraint on the unstable modes of the system leads to stability. They showed that existence of a solution

to the following optimization problem guarantees stability.

$$\min_{u_k} \sum_{i=1}^{m-1} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||\Delta u(k+i-1)||^2_{\Lambda_i} + ||x(k+m)||^2_{R_s} \quad (2.44)$$

$$\begin{aligned} A_u x(k+m) &= 0 \\ u_k &\in \Omega_k \end{aligned}$$

$A_s x$ and $A_u x$ are the projections of x onto the stable and unstable eigen spaces of A respectively. $R_s = A_s \bar{Q}_s A_s$, where $\bar{Q}_s$ is the positive definite solution to the Lyapunov equation

$$(A_s^\dagger)^T A^T A_s^T \bar{Q}_s A_s A A_s^\dagger - \bar{Q}_s + (A_s^\dagger)^T \bar{Q}_s A_s^\dagger = 0$$

with $A_s^\dagger = A_s^T (A_s A_s^T)^{-1}$.

The third approach consists of imposing a contraction constraint on the state of the system and was proposed by Polak and Yang (1993) . This is a weaker condition than the terminal state constraint and the feasible region for the inputs is not so restricted. Their algorithm can be summarized in the following optimization problem,

$$\min_{u_k} \sum_{i=1}^{p_k} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{p_k-1} ||\Delta u(k+i-1)||^2_{\Lambda_i} \quad (2.45)$$

$$\begin{aligned} ||x(k+p_k)||_R &\leq \alpha ||x(k)||_R \\ u_k &\in \Omega_k. \end{aligned}$$

where R is a positive definite matrix and $\alpha \in [0,1)$. Polak and Yang (1993) included the horizon length and the sampling time as optimization variables in their algorithm. Contrary to the earlier approaches the entire sequence of control moves is implemented before solving a new optimization problem. The contraction condition guarantees stability assuming that there exist feasible inputs.

The fourth and the last approach is to adopt the infinite prediction horizon and the infinite input horizon problem (Kothare *et al.*, 1996; Scokaert and Rawlings, 1998?). The optimization problem solved at every instant is:

$$\min_{u_k \in \Omega_k} \sum_{i=1}^{\infty} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{\infty} ||u(k+i-1)||^2_{\Lambda_i}. \quad (2.46)$$

This is equivalent to solving a constrained linear quadratic regulation (LQR) problem and subsequently an infinite dimensional optimization problem. Scokaert and Rawlings (1998) present an algorithm based on solution of a finite number of finite-dimensional convex

quadratic problems. Stability properties of the constrained LQR is again proven via Lyapunov arguments. Lee (1996) extends the stability results to the output feedback case based on Lipschitz continuity of the feedback law.

In a recent article, Lee *et. al* (1998) propose a matrix inequality conditions on the terminal weighting matrix under which closed loop stability is guaranteed for system with state and input constraints. The authors show that such a weighting matrix can be obtained by solving a linear matrix inequality (LMI).

2.6.2 Performance

Tracking

Traditionally tuning methods for MPC focused on heuristic selection of the tuning parameters- p, m, Γ, Λ. The formulation of the predictor in terms of the differenced input, Δu, results in the closed loop system being a type I system (Astrom and Wittenmark, 1991) *i.e.*, it can asymptotically track step changes in the setpoint. Thus conventional MPC is able to track step changes in the setpoint due to the presence of integral action,

$$u(k) = u(k-1) + \Delta u(k).$$

The speed of response, overshoot and other measures of tracking performance have to be optimized by careful selection of the tuning parameters. It is possible to impose constraints on the output to ensure minimal overshoot and speed of response, *etc.* Masuda and Shah (1998) exploit the two degree-of-freedom structure to design state space based GPC for better tracking. In other words the setpoint trajectory, $r(k+i)$, offers an additional degree of freedom. In general this degree of freedom can be utilized to optimize the closed loop tracking performance.

Disturbance Attenuation

Measured disturbances can be easily included in the predictive formulation. The predictor is suitably modified to include the effect of measured disturbances,

$$\hat{y}(k) = \sum_{j=1}^{N} h^u(j)u(k-j) + \sum_{j=1}^{N} h^d(j)d(k-j). \tag{2.47}$$

In the p-step ahead predictor the disturbance is assumed to be constant over the prediction horizon. The objective function remains unaffected except for the change in predictor. The input sequence is then calculated in such a manner so as to trade off the tracking disturbance rejection requirements optimally as per the MPC objective function. This

amounts to having a feedforward controller. The unmeasured disturbances are handled in different ways depending upon the predictor. The transfer function form has a noise model and incorporates the effect of unmeasured disturbances through a optimal p-step ahead predictor. The state space form uses the Kalman filter to include the effect of unmeasured disturbances. The OE form uses feedback to estimate the stochastic inputs and assumes that the disturbances will stay constant over the prediction horizon.

The success of MPC has led many researchers to extend the finite horizon idea to the state estimation problem (Muske *et al.*, 1993; Michalska and Mayne, 1992). A finite dimensional QP, which is a dual of the controller design optimization problem, is solved to compute the receding horizon estimator. Nominal stability of this observer can be proven on the basis of the stability of the Kalman filter. Furthermore the estimated states can be constrained to lie within a feasible region. This is an attractive feature since it can prevent infeasibility problems at the control stage.

In practice many disturbances are unknown in terms of their structure and effect on the controlled variables. This gives rise to uncertainty in the disturbance model. Design of MPC for robust performance with respect to uncertainty in disturbances is as yet an unresolved problem. The receding horizon nature of predictive controllers is one such feature which attempts to counter this kind of uncertainty.

2.7 Robust MPC

2.7.1 Modeling the uncertainty

The true nonlinear plant can be approximated by a linear model only up to a certain degree of accuracy. This leads to an inevitable model plant mismatch (MPM) which can be captured in several ways. The classical way of modeling this uncertainty consists of getting norm bounds on the uncertainty in the frequency domain (see Doyle *et al.*, 1992) . While this form of unstructured uncertainty is suitable for design and analysis of unconstrained controllers, the presence of constraints leads to formidable difficulties in utilizing the norm-bounded description for design and analysis of constrained MPC. As a result the following uncertainty descriptions have been found to be amenable for the design and analysis of constrained MPC.

1. **Interval uncertainty in the impulse response coefficients**. The true plant is represented by a family of impulse response models *i.e.* bounds on the impulse response coefficients are used to capture the uncertainty. This way of representing the

mismatch is control relevant since the FIR form is the most commonly used model form and the uncertainty is easily quantifiable,

$$P \in \wp \triangleq \{y(k) = \sum_{j=1}^{N} h(j)u(k-j) | h(j) \in [\underline{h} \leq h(j) \leq \bar{h}\}. \tag{2.48}$$

2. **Ellipsoidal bounds**. Instead of bounding each FIR coefficient by a lower and an upper bound, an ellipsoidal region quantifying the variability of the parameters can be naturally derived as an outcome of the identification process. If we let $\theta = [h(1).....h(N)]^T$ then the uncertainty is described by the set:

$$\theta \in \Theta \triangleq \{\theta | (\theta - \theta_0)^T W (\theta - \theta_0) \leq \epsilon\}. \tag{2.49}$$

where $W \succ 0$ is the parameter covariance matrrix and ϵ is a probability measure indicating the confidence level. Thus the uncertainty set is described by an ellipsoid centered at θ_0. The eigen structure of W determines the orientation and size of this uncertainty set. This type of uncertainty description captures the correlation structure of the parameters.

3. **A multi-model description**. Instead of relying on the parametrizing the process in terms of a large number of impulse response coefficients, a state space model with few parameters is used to represent the process. The model parameters belong to a convex set described by

$$x(k+1) = Ax(k) + Bu(k) \tag{2.50}$$
$$y(k) = Cx(k)$$
$$[A \;\; B] = \{\sum_{i=1}^{l} \lambda_i [A_i \;\; B_i] | \sum_{i=1}^{l} \lambda_i = 1, \lambda_i \geq 0\}.$$

Similarly the other parameters of the state space description belong to respective combinations of the multiple models (Kothare *et. al*, 1996).

Analysis of robustness properties has also been attempted by several researchers. Garcia and Morari (1982) looked at the robustness properties of unconstrained MPC in the IMC framework. Zafiriou (1990), Zafiriou and Marchal (1991) used the contraction mapping theorem to analyze the robustness properties of constrained MPC. Bannerjee and Shah (1995) used the Small Gain Theorem (SGT) to investigate the robustness of GPC type algorithms. Non-parametric modeling was used to estimate an upper bound on the additive

uncertainty. Genceli and Nikolaou (1993) used bounds on the impulse response coefficients to study the robustness of l_1 norm based MPC. They propose de-tuning the controller to improve the robustness of the controller. The input move suppression factor is increased to make the controller more passive and thereby stabilize the closed loop system.

Robust Design

Campo and Morari (1987) and Zheng and Morari (1993) present schemes for robust synthesis of MPC based on the min-max principle (minimizing the worst case value of the objective function, where the worst case is taken over the set of uncertain plants). They use uncertain impulse response coefficients for describing the uncertainty in the system. The following min-max problem is derived from game theory,

$$\min_{u_k \in \Omega_k} \max_{P \in P} \sum_{i=1}^{p} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||u(k+i-1)||^2_{R_i}. \tag{2.51}$$

In the case of a FIR model the optimization problem can be shown to be convex. Very few stability properties of this feedback law have been proven (Zheng and Morari, 1993). Furthermore this approach cannot be extended to the case of unstable plants.

Kothare *et al.* (1996) give an LMI framework for design of robust constrained MPC. They derive control laws based on the multi-model description of uncertainty and also for the case of structured feedback uncertainty. An infinite dimensional optimization problem is solved at every instant by first deriving an upper bound on the following objective function and then minimizing this upper bound,

$$\min_{u_k \in \Omega_k} \max_{P \in P} \sum_{i=1}^{\infty} ||x(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{\infty} ||u(k+i-1)||^2_{R_i}. \tag{2.52}$$

The robust stability of the subsequent feedback law is proven by showing that the objective function obtained from the optimal solution at time k, is a strictly decreasing Lyapunov function. The issue of robust performance is also addressed by Kothare and co-authors.

Ralhan (1998) uses the ellipsoidal description of uncertainty to design robust MPC. Robustness is ensured through addition of a quadratic constraint that prevents the achieved cost on the true plant being more than the design cost. A cost function based on the nominal model is minimized subject to this robustness constraint,

$$\hat{J}_k(u_k, \hat{\theta}) = \sum_{i=1}^{p} ||y(k+i)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||u(k+i-1)||^2_{R_i} \tag{2.53}$$

$$\min_{u_k \in \Omega_k} \hat{J}_k(u_k, \hat{\theta}) \tag{2.54}$$

$$\text{subject to } \hat{J}_k(u_k, \hat{\theta}) < J_k(\hat{u}_k, \theta).$$

where $\hat{u}_k$ is defined as the restriction of the optimal input sequence at the past sampling instant, $k-1$. The robustness constraint can be expressed as a quadratic constraint on the decision variables, u_k.

Recent developments in adaptive and nonlinear MPC are not covered here. Lee (1996) gives an overview of the latest developments in these areas. The area of nonlinear MPC is reviewed at a later stage. The review articles by Garcia *et al.* (1989) and Lee (1996) give an in depth discussion of the issues presented in this chapter. To summarize the material covered in this chapter:

- MPC is a very flexible controller. A variety of constraints can be accommodated in the predictive control formulation and this remains its greatest asset.
- A variety of model forms are compatible with MPC. While the FIR, albeit with the aid of a large number of parameters, can represent unusual high order behaviour, the state space and the transfer function models are parsimonious and therefore can be estimated more accurately.
- A large number of tuning parameters need to be carefully chosen in order to design a satisfactorily working MPC. There are a number of heuristic guidelines available but no consistent mathematical procedure to enable the optimal choice of these tuning parameters. One approach considered by Bannerjee (1996) was to formulate a mixed integer nonlinear program (MINLP) to compute the appropriate values of the prediction and control horizons and the weighting matrices.
- Recent trends in literature reveal a surge in the infinite horizon formulation owing to the ease of analyzing and designing this type of predictive control law. Use of Lyapunov theory appears to be the method of choice for stability analysis of constrained MPC.
- A number of methods for designing robust MPC are emerging. The multiple model and the ellipsoidal representation of the model plant mismatch are elegant ways of capturing the modeling uncertainty.

In the next chapter we will look at the use of linear matrix inequalities and convex optimization in the design of predictive controllers.

Chapter 3

Model Predictive Control via Linear Matrix Inequalities

3.1 Introduction

One of the computational steps in MPC involves solution of a constrained optimization problem. Depending upon the choice of the objective function, either a linear program or a QP arises at every sampling instant. The numerical techniques such as the simplex method or algorithms for solving QPs are exponential time algorithms *i.e.* the computational time increases as an exponential function of the problem dimensionality. A class of numerical methods known as the interior point methods were first shown to be polynomial time algorithms for linear programs by Karmarkar (1984). This led to further development and generalization of the interior point methods to the general class of convex problems. The linear matrix inequality (LMI) framework gives a way of casting quadratic and linear type of convex problems in terms of matrix variables (Boyd *et al.*, 1994).

In this chapter we give a novel way of casting the constrained MPC scheme as a semidefinite program (SDP) which belongs to a particular class of convex problems. The SDP formulation of the optimization step in MPC is outlined in a tutorial manner. Constraints on the inputs and the outputs are also incorporated. The LMI framework is also exploited to cast the selection of weighting matrices for MPC as a generalized eigenvalue problem (GEVP).

This chapter is organized as follows: Section 3.2 gives the background material on semidefinite programming. A brief overview of the applications of LMIs in process control is presented in Section 3.3. In Section 3.4 we discuss the formulation of the model predictive

[1] A version of this chapter has been published as "Constrained Long Range Predictive Control : Exploiting Convexity ", 1997, R. S. Patwardhan and B. Huang, In Proceedings of IFAC Symp. Adv. Control. Chem. Proc., pp. 317-322, Banff, Canada

controller design as a semidefinite program for the SISO case. A novel method for choosing the weighting matrices in the MPC objective function via the LMI framework is presented in Section 3.5. The results of the real-time applications are presented and discussed in Section 3.6, followed by concluding remarks in Section 3.7.

3.2 Semidefinite Programming

Convex problems have unique solutions. This makes them fundamentally tractable, both in theory and practice. One of their desirable properties is that the locally optimal solutions are globally optimal. SDPs are a class of convex optimization problems where the objective function is linear (Boyd and Vandenberghe, 1994) :

$$\begin{aligned} & \min\ c^T x \\ s.t.\ F(x) &\succeq 0 \\ \text{where } F(x) &= F_0 + x_1 F_1 + \ldots + x_m F_m \end{aligned} \tag{3.1}$$

$c \in R^m$ and $m+1$ symmetric matrices $F_0, .., F_m \in R^{m \times m}$. The inequality sign $F(x) \succeq 0$ implies $F(x)$ is positive semidefinite, *i. e.*, $v^T F(x) v \geq 0$ for all nonzero $v \in R^m$. In addition, the leading principal minors of $F(x)$ must be positive. The constraint $F(x) \succeq 0$ also defines a LMI in the nonstrict sense. A strict LMI implies $F(x) \succ 0$ (Boyd *et al.* , 1994) .

Convex quadratic constraints can be expressed as LMIs using the Schur complement result:

$$\begin{bmatrix} Q(x) & S(x) \\ S(x)^T & R(x) \end{bmatrix} \succ 0 \tag{3.2}$$

which is equivalent to

$$R(x) \succ 0, Q(x) - S(x) R(x)^{-1} S(x)^T \succ 0 \tag{3.3}$$

or

$$Q(x) \succ 0, R(x) - S(x)^T Q(x)^{-1} S(x) \succ 0$$

where $Q(x) = Q(x)^T, R(x) = R(x)^T$, and $S(x)$ depend affinely on x. For the non-strict case Boyd *et al.* (1994) give the following result

$$\begin{bmatrix} Q(x) & S(x) \\ S(x)^T & R(x) \end{bmatrix} \succeq 0 \tag{3.4}$$

which is equivalent to

$$R(x) \succeq 0, Q(x) - S(x)R(x)^{\dagger}S(x)^{T} \succeq 0, S(x)(I - R(x)R^{\dagger}(x)) = 0 \tag{3.5}$$

where † denotes the pseudo-inverse.

Other instances of convex problems include linearly constrained quadratic programs (LCQP) and quadratically constrained quadratic programs. A linear program (LP) is a special case of a semidefinite program as the example from (Boyd and Vandenberghe, 1994) illustrates,

$$\begin{aligned} & \min \ c^{T}x \\ s.t. \ Ax + b \ & \geq \ 0. \end{aligned} \tag{3.6}$$

This LP can be expressed as a SDP with $F(x) = diag(Ax + b)$, *i.e.*,

$$F_0 = diag(b), F_i = diag(a_i), i = 1, \ldots, m$$

where a_i is the ith column of A.

Semidefinite programs can be solved very efficiently using interior point methods. Interior point methods are polynomial time algorithms for convex problems. They were first developed by Karmarkar (1984) for efficient solution of LPs. Nesterov and Nemirovsky (1994) were the first to generalize the interior point methods to general convex programming. Wright (1997) gives a lucid description of the primal-dual class of interior point methods for a class of convex problems.

Wright (1996) proposed an interior point method that exploits the structure of the quadratic programming problem (QP) in MPC. An interior point based primal-dual approach to solving the QP is facilitated by rearranging the optimization variables. This results in a banded tridiagonal structure for the linear equations to be solved at each iteration of the interior point method. The inherent sparsity of the problem is thus exploited. Wright (1996) also proposes factorization techniques for updating the gradient information from one sampling instant to the next in an efficient way.

3.3 LMIs in Process Control

The use of linear matrix inequalities as tools for model predictive control has been increasing over the past few years. In this Section we review the applications of LMIs to process control in general and MPC in particular. Kothare *et al.* (1996) use LMIs to solve the robust

model predictive control problem. They derive an upper bound on the worst case infinite horizon objective function and then minimize this upper bound via LMIs and semidefinite programming to derive a time-varying state feedback law. The formulation of Kothare *et al.* is quite general and includes contraction constraints on the states for speeding up convergence. Conventional input and output constraints are handled by obtaining sufficient LMI constraints. Robust stability is proved via Lyapunov arguments on the upper bound.

Genceli and Nikolaou (1996) proposed a novel approach to adaptive MPC which designed an input sequence that simultaneously regulated and excited the plant. An additional persistent excitation constraint was imposed on the optimization step in MPC,

$$\min_{u_k \in \Omega_k} J_k(u_k) \tag{3.7}$$

$$\text{Persistant Excitation Constraint} \;:\; \sum_{j=0}^{m-1} \lambda^j \phi(k-j+i)\phi(k-j+i)^T \succeq (\rho_0 - \mu)I$$

where $\phi(k) = [u(k-1)\ u(k-2)....u(k-N)\ 1]^T$. The persistent excitation constraint is a non-convex quadratic matrix inequality(QMI). Genceli and Nikolaou approximate this QMI with an LMI and use an ellipsoid algorithm to solve the resulting SDPs iteratively till convergence is achieved.

Lee *et al.* (1998) propose a matrix inequality constraint on terminal weighting matrix and show the stabilizing properties of the resulting constrained receding horizon law for time-varying discrete linear systems. The following objective function is minimized,

$$J_k = \sum_{i=1}^{p-1} (||x(k+i)||^2_{\Gamma_i} + ||u(k+i)||^2_{\Lambda_i}) + ||x(k+p)||^2_{\Psi_k}$$

where the terminal state weighting matrix satisfies the matrix difference inequality

$$\Psi_k \succeq F_k^T \Psi_{k+1} F_k + \Gamma_k + H_k^T \Lambda_k H_k. \tag{3.8}$$

$H_k = A + BH_k, H_k \in R^{n_u \times n_y}$. This matrix inequality is then expressed as an LMI in terms of appropriate matrix variables. For the constrained case an artificial invariant ellipsoid constraint, which is a convex quadratic constraint, is introduced to relax the terminal state constraint.

In a novel application of LMIs to a process control application, Thake *et al.* (1997) used the SDP formulation to approximate a high dimensional multivariable controller using a reduced order controller with special structure. The approximation problem was reduced to a spectral norm minimization problem of the form

$$\min_{X_{i,j}} ||X - K||_2 \tag{3.9}$$

where K is the high dimensional controller and X has far fewer elements, e.g., a banded tridiagonal structure. The approximation technique was demonstrated on paper machine cross-direction (CD) profile control problem.

Several quadratic constraints considered in Chapter 2 can be expressed as LMIs. For example the contraction constraint can be expressed as an LMI

$$||x(k+p)||_2 \leq \alpha ||x(k+1)||_2 \Leftrightarrow$$

$$\Delta u_k^T S_p^T S_p \Delta u_k + \Delta u_k^T S_p^T F_p x(k) + c(k) \leq 0)$$

where $c(k) = x(k) F_p^T F_p x(k) - \alpha x(k)^T x(k)$. This is equivalent to the following LMI,

$$\begin{bmatrix} I & \Delta u_k S_p \\ S_p^T \Delta u_k^T & \Delta u_k^T S_p^T F_p x(k) + c(k) \end{bmatrix} \succeq 0. \tag{3.10}$$

The robustness constraint of Ralhan (1998) can be expressed as an LMI,

$$\hat{J}_k(u_k, \hat{\theta}) < J_k(\hat{u}_k, \theta) \Leftrightarrow \tag{3.11}$$

$$\begin{aligned} \max_{\theta \in \Theta} (u_k^T u_k - \hat{u}_k^T \hat{u}_k + \lambda \theta^T (U_k^T U_k - \hat{U}_k^T \hat{U}_k) \theta) &\leq 0 \Leftrightarrow \\ u_k^T u_k - \hat{u}_k^T \hat{u}_k + \lambda \theta^{*T} (U_k^T U_k - \hat{U}_k^T \hat{U}_k) \theta^* &\leq 0 \end{aligned}$$

where $U_k, \hat{U}_k$ and the derivation of θ^* are given in Ralhan (1998).

3.4 A SDP Formulation of model Predictive Control

3.4.1 The unconstrained case

In this Section we discuss the steps involved in posing the MPC design procedure as a convex problem. The MPC objective function can be rewritten as:

$$J = (r_k - f_k - d_k)^T \Gamma (r_k - f_k - d_k) + \Delta u_k^T (S^T \Gamma S + \Lambda) \Delta u_k - 2 \Delta u_k^T S^T \Gamma S (r_k - f_k - d_k).$$

We can ignore the constant term in the objective function, $(r_k - f_k - d_k)^T \Gamma (r_k - f_k - d_k)$. We then define a scalar γ such that

$$\Delta u_k^T (S^T \Gamma S + \Lambda) \Delta u_k - 2 \Delta u_k^T S^T \Gamma S (r_k - f_k - d_k) \leq \gamma.$$

Having introduced an auxiliary variable that serves as an upper bound on the objective, we can recast our optimization problem as

$$\begin{aligned} &\min \ \gamma \\ &\text{subject to} \\ &\Delta u_k^T (S^T \Gamma S + \Lambda) \Delta u_k - 2 \Delta u_k^T S^T \Gamma S (r_k - f_k - d_k) \leq \gamma. \end{aligned}$$

In this formulation the objective function is linear with respect to ϕ; the earlier quadratic objective function appears as a quadratic constraint. This quadratic constraint can be expressed in terms of a LMI using the Schur Complement for the non-strict case (Vandenberghe and Boyd, 1996) ,

$$\begin{array}{l} \min\ \gamma \\ \text{subject to} \\ \begin{bmatrix} I & \Delta u_k L \\ L^T \Delta u_k^T & 2(r_k - f_k - d_k)^T S \Delta u_k + \gamma \end{bmatrix} \preceq 0 \end{array} \tag{3.12}$$

where $L^T L$ is the Cholesky Factorization of $(S^T \Gamma S + \Lambda)$. This is the standard form of a SDP.

3.4.2 Extension to the constrained case

Usually a real process involves rate and amplitude constraints on the input, and may also require output constraints to be considered. As per Section 2.5, the constraints can be represented by the linear inequality,

$$A\Delta u_k + B_k \geq 0$$

which is equivalent to the following LMI,

$$diag(A\Delta u_k + B_k) \succeq 0. \tag{3.13}$$

Incorporating equation 2.33 as an additional LMI, the earlier SDP (equation 3.12) now becomes

$$\begin{array}{l} \min\ \gamma \\ \text{subject to} \\ \begin{bmatrix} I & \Delta u_k L \\ L^T \Delta u_k^T & 2(r_k - f_k - d_k)^T S \Delta u_k + \gamma \end{bmatrix} \preceq 0 \\ diag(A\Delta u_k + B_k) \succeq 0 \end{array} \tag{3.14}$$

This formulation also belongs to the class of SDPs. Thus we have posed a linearly constrained quadratic program as a SDP. The MIMO case can be formulated as a SDP in a similar manner. The case study discussed in Section 3.6.2 is an application of the MIMO approach. The SDP approach can also be extended to receding horizon techniques based on state space or impulse response models, *etc.* The convexity of the underlying optimization problem is independent of the model structure and the type of disturbance models, provided they are linear models.

3.5 Choice of weighting matrices via LMIs

There are no general guidelines for choosing weighting matrices in the MPC objective function. Heuristic rules are often used to choose initial weighting matrices. In the ensuing discussion we illustrate a systematic way of choosing the output and input weighting matrices - Γ, Λ, so as to minimize the condition number of $S^T\Gamma S + \Lambda$. Consider the example in Section 2.5 with $p = 6, m = 2$, the dynamic matrix is given by:

$$S = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0.0778 & 0 \\ 0.1495 & 0.0778 \\ 0.2158 & 0.1495 \\ 0.2769 & 0.2158 \end{bmatrix}$$

$$cond(S) \triangleq \frac{\sigma_{\max}(S)}{\sigma_{\min}(S)} = 79.9755$$

We solve the following generalized eigenvalue problem (Boyd *et al.* , 1994), to find out the weightings Γ, Λ to minimize the condition number of $S^T\Gamma S + \Lambda$,

$$\begin{aligned} & \min_{\Gamma,\Lambda} \gamma^2 \\ & \text{subject to} \\ \Gamma &\in R^{n_y p \times n_y p} \text{ and diagonal, } \Gamma > 0 \\ \Lambda &\in R^{n_u m \times n_u m} \text{ and diagonal, } \Lambda > 0 \\ I &\leq S^T\Gamma S + \Lambda \leq \gamma^2 I. \end{aligned} \tag{3.15}$$

The LMI toolbox in Matlab was used to solve the above GEVP. The following weighting matrices gave a condition number of 1.0002 for $S^T\Gamma S + \Lambda$,

$$\begin{aligned} \Gamma &= diag(1.5735, 1.5735, 1.5766, 0.0070, 0.0025, 0.0014) \\ \Lambda &= diag(1.9901, 1.9998). \end{aligned}$$

This is an impractical choice of Γ since the last three predictions have essential zero weights. As a remedy, we constrained the choice of elements in Γ, Λ to the interval [0,10]. The GEVP in equation 3.15 is solved subject to the additional constraints:

$$\begin{aligned} 0 &\leq \Gamma_i \leq 10 \\ 0 &\leq \Lambda_i \leq 10. \end{aligned}$$

The following Γ, Λ gave a condition number of 1.0355:

$$\begin{aligned}\Gamma &= diag(5.5000, 5.5000, 5.5125, 1.5224, 1.1889, 1.1025)\\ \Lambda &= diag(7.2498, 7.2839).\end{aligned}$$

This presents a systematic way of choosing appropriate weighting matrices for MPC. An arbitrary choice of $\Gamma = I, \Lambda = 0.1I$ gives $cond(S^T\Gamma S + \Lambda) = 3.15$. The stability results reviewed in last chapter can then be used to a stabilizing feedback MPC law.

3.6 Real-time Control using LMIs

The SDP formulation described above was evaluated experimentally on two computer-interfaced pilot-scale processes. A MPC type controller was developed and the optimization step was solved using a semidefinite programming solver, based on the projective method of Nesterov and Nemirovsky (1994) available as part of the LMI Toolkit Gahinet *et al.*, 1995. In all the cases, the LMIs were set up in Matlab 4.0 using the LMI Toolkit.

3.6.1 SISO Process

The constrained MPC algorithm using the LMI toolbox was evaluated on a computer interfaced temperature control process. A schematic of the process is shown in figure 3.1. The objective is to control the temperature in a region around the light bulb by manipulating the voltage across the bulb filament. A constant speed fan provides the cooling action and also serves as the source of a disturbance. A thermocouple measures the temperature in the region of interest. The transfer function between the power(input) and temperature(output) was first identified as:

$$\frac{Y(z)}{U(z)} = \frac{0.0173z^{-3} + 0.0047z^{-4} + 0.0069z^{-5}}{1 - 0.8484z^{-1} + 0.08z^{-2} - 0.1484z^{-3}}. \tag{3.16}$$

The real-time implementation was carried out using Realtime-Simulink workshop. The sampling time was 5 seconds. The controller tuning parameters selected were $p = 20, m = 2, \Lambda = 6I$. The following rate and amplitude constraints were imposed on the input: $0 \leq u(k) \leq 100$, $-20 \leq \Delta u(k) \leq 20$. The nominal operating conditions were $u_{ss} = 55\%; y_{ss} = 59.5\%$.

Figure 3.2 shows the tracking performance and the disturbance rejection characteristics of the designed controller. Rate and amplitude constraints on the input were satisfied (see figure 3.3). At the time instant 100, when the setpoint is changed by a large value from 60 to 80, the input saturates at its high limit and as expected the output cannot reach the

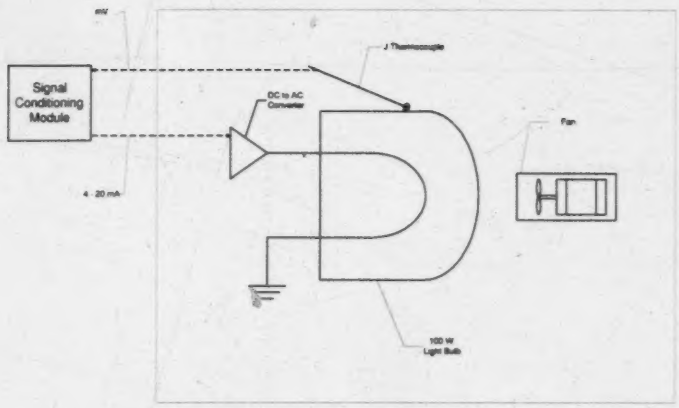


Figure 3.1: *Schematic of the light bulb experiment.*

desired value and a bias in the control error is observed. The same phenomenon is observed when another large setpoint change is effected at sampling instant 150.

3.6.2 MIMO Process

The SDP formulation of the model predictive controller was tested on a temperature-level control of a mixing process. Cold water (5-10 °C) and hot water (45-50 °C) flow through two inlet pipes into a glass tank. The exit flow rate of water from the glass tank is not controlled. The control objective was to maintain the temperature (y_1) and level (y_2) of Tank 1 by manipulating the inlet hot water (u_1) and cold water (u_2) flow rates. The input-output relations were identified using a multivariable identification routine (Badmus *et al.* ,1996) based on canonical variate analysis (CVA). The transfer function matrix as identified is:

$$\hat{P}(z^{-1}) = \begin{bmatrix} \frac{0.0235z^{-1}}{1-0.8607z^{-1}} & \frac{-0.1602z^{-1}}{1-0.8607z^{-1}} \\ \frac{0.2043z^{-1}}{1-0.9827z^{-1}} & \frac{0.2839z^{-1}}{1-0.9827z^{-1}} \end{bmatrix}. \tag{3.17}$$

The multivariable MPC was implemented in a Realtime-Simulink environment. The sampling time was chosen as 5 seconds.

The runs carried out demonstrate the applicability of the LMI based MPC. The MPC tuning parameters were $p = 10, m = 2, \Lambda = 0.1I$. The rate and amplitude constraints were imposed on both inputs were respectively $0 \leq u_i(k) \leq 100$ and $-20 \leq \Delta u_i(k) \leq 20, i = 1, 2$.

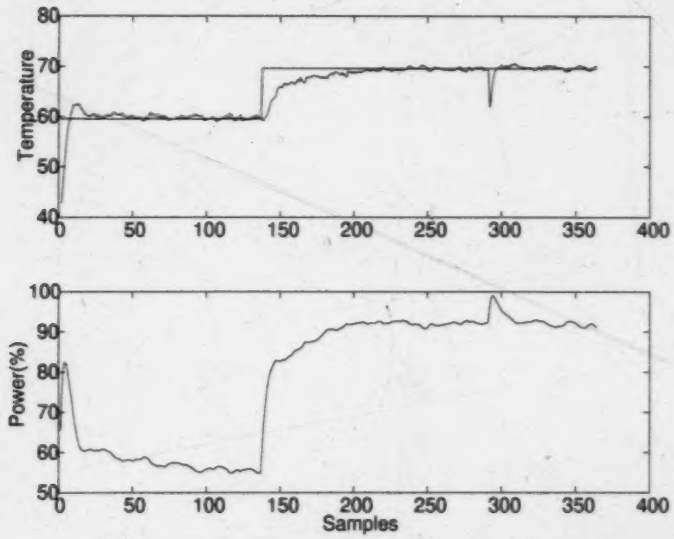


Figure 3.2: *Performance of constrained LMI based MPC for the temperature control of a light bulb process.*

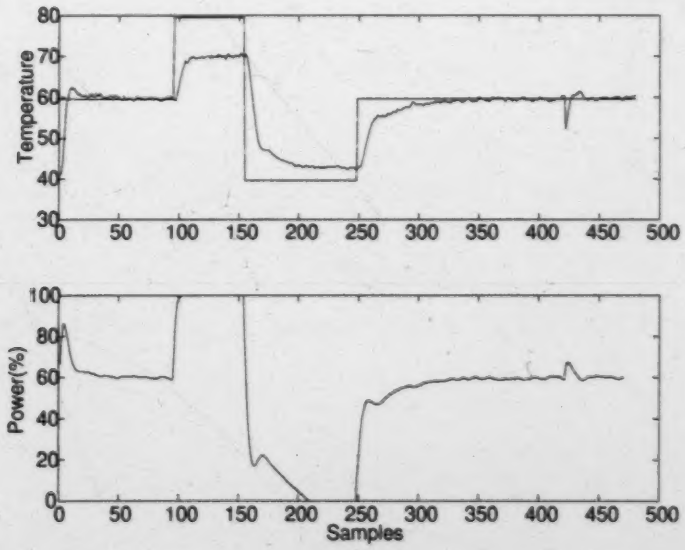


Figure 3.3: *Temperature control of a light bulb process: Effect of constraints.*

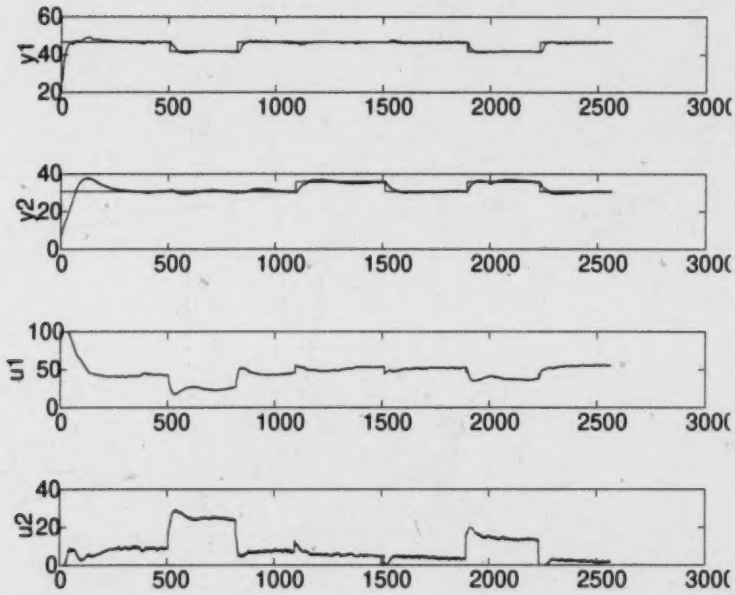


Figure 3.4: *Tracking performance of the constrained MPC for the temperature and level control of a mixing process*

The steady state operating conditions were:

$$u_{1ss} = 29\%; u_{2ss} = 16\%; y_{1ss} = 46.7^{o}C, y_{1ss} = 16\ cm.$$

The tracking and regulatory characteristics of the designed controller are shown in figs. 3.4 and 3.5, respectively. Rate and amplitude constraints on the input were satisfied during the tracking run. Notice the instances in figure 3.4 when the inputs saturate at the above specified limits. This illustrates that the algorithm indeed does enforce the specified limits. The hot water inlet temperature showed a time-varying trend. Figure 3.5 illustrates the ability of the controller to compensate for this time-varying disturbance and maintain the outputs at their desired values.

We compared the computational effort required by the quadratic programming solver *qp* and the semidefinite programming solver *mincx*. Both of these above solvers are available as functions in the optimization and LMI toolboxes in Matlab® respectively. Table 3.1 presents the computational times (CPU time in seconds) and floating point operations (flops), required by the two solvers for calculating the entire sequence of control moves for the simulation case study in Section 3.4. The dimensionality (m) of the optimization problem remains the same but the nature and number of constraints changes from case 1

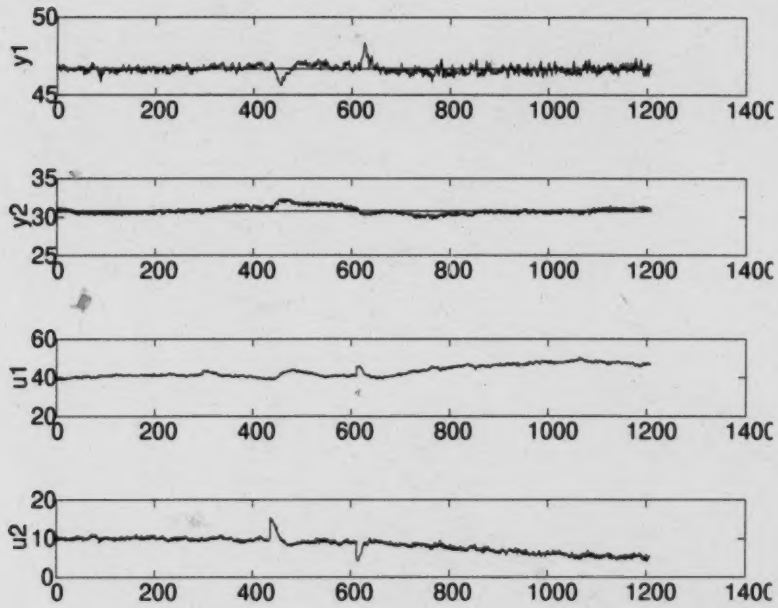


Figure 3.5: *Controller performance in the regulatory mode.*

to case 3.

- Case 1: Rate Constrained MPC ($2m$ constraints)
- Case 2: Amplitude Constrained MPC ($2m$ constraints)
- Case 3: Rate plus Amplitude Constrained MPC ($4m$ constraints)

The semidefinite programming solver proves to be an order of magnitude slower than the quadratic programming solver. The computational times for the mixing process (2×2) case were compared with respect to changing dimensionality ($2m$). Table 3.2 reports the CPU times/flops for computing a single control move for each solver. In this case the dimensionality of both, the problem size and the constraint set increase with increasing m. In this case the CPU times of the SDP solver increased rapidly whereas the QP solver showed stable behavior. The floating point operations, however, were comparable. The quadratic programming solver *qp* proved to be far more efficient than the convex programming solver *mincx*. The ineffectiveness of *mincx* could be attributed to its inefficient handling of the canonical form.

3.7 Conclusions

The MPC design problem is cast as a semidefinite program and solved using a SDP method. The proposed approach is demonstrated on SISO and MIMO processes- in a real-time environment. To the best of our knowledge this is the first experimental evaluation of the constrained MPC based on the convex optimization framework using interior point methods. A systematic technique of choosing the weighting matrices based on minimizing the condition number of a suitable matrix, via the LMI framework was also presented.

The following additional remarks are based on our experience with this method:

- The efficiency of the interior point methods can be better exploited for the multivariate case, where large dimensionality and structure play an important role.
- The predictive control problem occurs naturally as a linearly constrained quadratic program. Interior point methods developed for LCQPs may prove to be superior for this case, *e.g.* Wright (1996).
- It should also be noted that the LMI control toolbox is optimized for block structured LMIs with matrix variables. The toolbox is quite inefficient in handling inequalities in the canonical form. This results in higher computational times as compared to standard optimization software for solving QPs. This fact is also corroborated by Kothare et al. (1996) .
- The performance of convex optimization methods vis-à-vis local methods needs to be assessed. There is a need to quantify the benefits of using the above formulation.

Inspite of these shortcomings we would like to emphasize here that interior point methods provide tools for developing very reliable design procedures. Convergence to the global minima is guaranteed and interior point methods have polynomial complexity in the worst case. As more efficient implementations are developed, the software will be able to match the efficiency promised by these methods[1].

[1]This was attempted in 1995-96. Currently available methods and software is more efficient.

Table 3.1: *Computational Effort for simulation casestudy (m=2)*

	CPU time (flops)	
	qp	*mincx*
Case 1	2.69(101385)	19.83(167312)
Case 2	2.59(104369)	20.05(163712)
Case 3	2.80(112985)	23.56(195740)

Table 3.2: *Computational Effort for the MIMO case*

	CPU time (flops)	
	qp	*mincx*
$m = 2$	0.11(43184)	2.36(48784)
$m = 3$	0.11(48064)	9.44(61912)
$m = 4$	0.11(56179)	28.3(85316)
$m = 5$	0.11(132350)	77.2(122090)

Chapter 4

A Partial Least Squares Framework for Model Predictive Control

4.1 Introduction to Modeling and Control via Partial Least Squares

Partial least squares (PLS) has established itself as a robust alternative to the standard linear least squares technique in the analysis of correlated data. First proposed by Wold (1966), this technique has found application in many disciplines such as social sciences, engineering, chemistry and medicine. In the field of chemical engineering, PLS has found applications in the area of process monitoring, modeling, fault detection and control.

In PLS, the input (U) and output (Y) data is expressed as a sum of rank one outer products,

$$\begin{aligned} U &= t_1 p_1^T + t_2 p_2^T + \ldots + t_n p_n^T + E_{n+1} = TP^T + E_{n+1} \\ Y &= v_1 q_1^T + v_2 q_2^T + \ldots + v_n q_n^T + F_{n+1} = VQ^T + F_{n+1}. \end{aligned} \tag{4.1}$$

where n is the number of PLS dimensions, E_{n+1}, F_{n+1} are the residual portions of the inputs and the output data respectively. In the above representation, T and V denote the scores, henceforth referred to as the latent variables (LVs). P and Q represent the loading matrices for the U and Y blocks respectively. The first set of loading vectors p_1 and q_1, is obtained by maximizing the covariance between U and Y. Projection of the U and Y data onto the

[1]Versions of some sections of this chapter have been presented or appeared as: (a) R. S. Patwardhan, S. Laskhminarayanan and S. L. Shah, 1997, "Interactor Estimation and Multivariable Time Delay Compensation in the Partial Least Squares Framework", AIChE Annual Meeting, LA. (b) S. Lakshminarayanan, R. S. Patwardhan, S. L. Shah and K. Nandakumar, 1997, "A Dynamic PLS Framework for Constrained Model Predictive Control", In Procedings of IFAC Symp. Adv. Control. Chem. Proc., Banff, Canada. and (c) R. S. Patwardhan, S.Lakshminarayanan and S. L. Shah, 1998, "Constrained Nonlinear MPC using Hammerstein and Wiener Models- A PLS Framework", *AIChEJ.*, 44(7), 1611-1622.

loading vectors gives the first set of scores

$$\begin{aligned} t_1 &= Uj_1 \\ v_1 &= Yq_1 \end{aligned} \tag{4.2}$$

where $j_1 = \frac{U^T t_1}{t_1^T t_1}$. A linear regression between the scores results in the following *inner* relationship,

$$\widehat{v}_1 = t_1 b_1. \tag{4.3}$$

In equation 4.3 $\hat{v}_1 q_1^T$ may be interpreted as the part of Y data that has been predicted by the first PLS dimension; in doing so, the $t_1 p_1^T$ portion of U has been utilized. The above procedure is repeated on appropriate residuals from the U and Y blocks to find the remaining scores/loading vectors till a major portion of the covariance between U and Y has been explained. Figure 4.2 illustrates this iterative procedure. This procedure is known as the PLS-NIPALS algorithm and was first developed by Wold (1966). The PLS technique has also been cast into the Singular Value Decomposition (SVD) framework (Wise, 1991).

From a practical point of view, PLS can be considered as a technique that decomposes a multivariate regression problem into a series of univariate regression problems. The PLS based regression model can be written as

$$\begin{aligned} \widehat{Y} &= \widehat{V}Q^T + F_{n+1} = TBQ^T + F_{n+1} \\ \text{where } B &= diag(b_1, b_2, \ldots, b_n). \end{aligned} \tag{4.4}$$

Here the b_i's refer to the coefficients of the inner SISO linear regression models.

4.1.1 PLS based Dynamic Models

If the U block were to include the past values of the inputs/outputs, then analysis based on PLS would yield a dynamic model of the form

$$Y = UC_{dyn} + e. \tag{4.5}$$

With this approach , C_{dyn} can be interpreted as a matrix, whose elements are the finite impulse response coefficients (Ricker, 1988) or multivariable autoregressive moving average (ARMA) model terms (Qin and McAvoy, 1992). Kaspar and Ray (1992, 1993) developed PLS based dynamic models based on filtering of input data. The major dynamic component in the data was explained by filtering the data and then the standard PLS procedure is used to model the steady state characteristics from the filtered data.

Lakshminarayanan *et al.* (1997a) proposed a dynamic extension based on a modification of the PLS inner relation. Instead of relating the input and output scores (t_i and v_i) using a static model, a dynamic component such as ARX is used. This gives the dynamic analog of equation 4.4 as:

$$\hat{Y} = G_1(t_1)q_1^T + G_2(t_2)q_2^T + \ldots + G_n(t_n)q_n^T + F_{n+1}. \tag{4.6}$$

Here G_i's denote the linear dynamic models (e.g. ARX) identified at each stage. In terms of the latent variables we can write

$$V(z^{-1}) = \begin{bmatrix} G_1(z^{-1}) & 0 & \cdots & 0 \\ 0 & G_2(z^{-1}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & G_n(z^{-1}) \end{bmatrix} T(z^{-1}) = G(z^{-1})T(z^{-1}). \tag{4.7}$$

As per figure 4.1 the relationship between the original variables is given by:

$$Y(z^{-1}) = \{QG(z^{-1})P^{\dagger}\}U(z^{-1}). \tag{4.8}$$

Note that G is a diagonal transfer function matrix (inner relationship), but P and Q are full

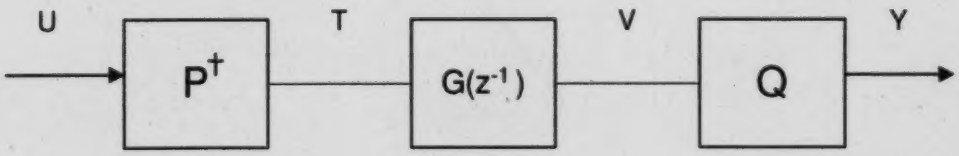


Figure 4.1: *Partial Least Squares based Dynamic Model*

matrices (loadings). As a result the overall relationship between U and Y is a full transfer function matrix given by:

$$G_o(z^{-1}) = QG(z^{-1})P^{\dagger}. \tag{4.9}$$

Thus, despite the diagonal nature of the inner relationship, the overall model in terms of the original variables is multivariable and does capture the interactions amongst different inputs and outputs.

Remark 4.1. *The dynamic PLS models are particularly well suited for controller design, since in terms of the latent variable framework we have several univariate models which can be used to design controllers based on SISO theory. This presents a strong case for*

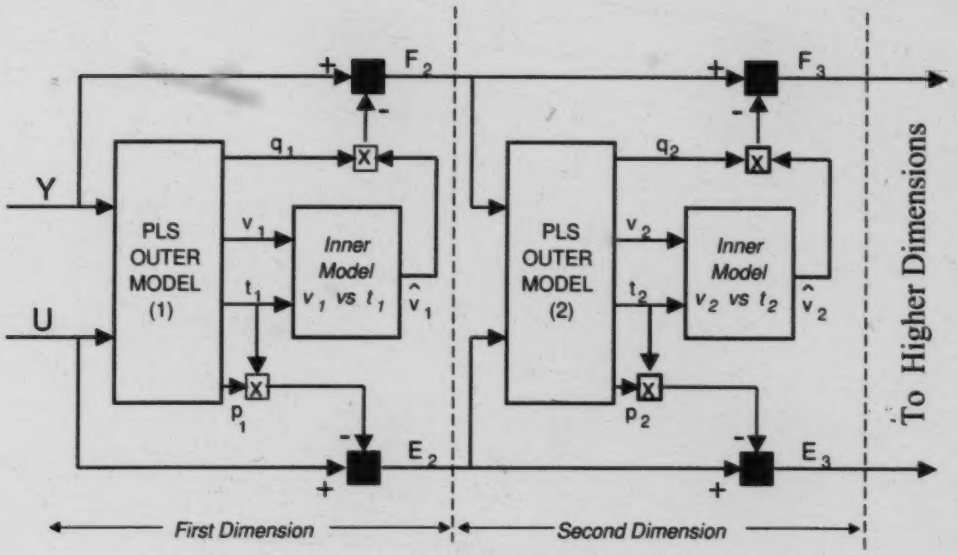


Figure 4.2: *A schematic diagram showing the PLS modeling procedure.*

the design of controllers in latent space. Suitable pre- and post-compensators have to be used to transform the variables from the latent to the original space and vice-versa. For the unconstrained case, the MIMO controller design problem is thus reduced to several SISO design problems.

Remark 4.2. *The PLS estimation procedure is nonlinear by itself. Hence accuracy of parameter estimates, variance of estimates,* etc., *is not easy to quantify. The goodness of a PLS model has to be often determined through cross-validation and residual analysis alone.*

The decoupled nature of the PLS models leads to a very intuitive interpretation of the *interactor matrix*, which is the multivariate generalization of the time delay term as in the univariate case. It plays a crucial role in controller design for linear, discrete systems. In the latent space the delays in each channel can be factored out to yield a diagonal interactor. Transforming back to the original domain gives the interactor matrix in the original or the Euclidean basis space. This is a simple method of estimating the interactor matrix. Delays can often be troublesome in controller design. For the MIMO case prior knowledge of the interactor matrix is an asset in dead time compensation. When the controller design is done using the PLS based dynamic models, Smith predictors can be designed for each PLS-based controller and the overall effect is equivalent to that of a multivariable time delay compensator based on the interactor factorization.

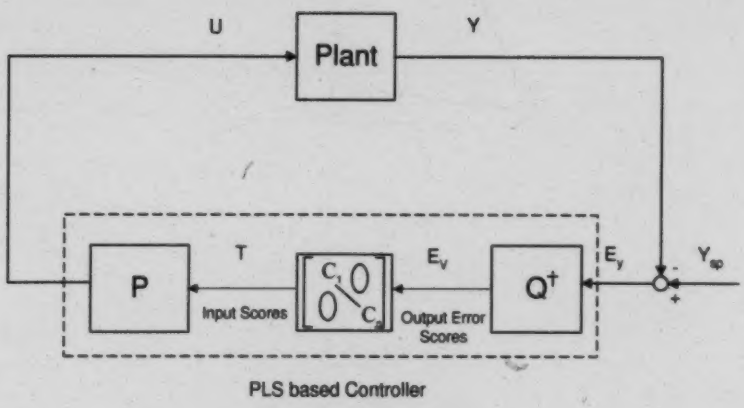


Figure 4.3: *Controller synthesis in the latent space framework*

4.1.2 Controller Synthesis in the PLS Framework

The dynamic PLS models are particularly well suited for controller design, since in terms of the latent variables we have several univariate models (equation 4.7) which can be used to design controllers based on SISO theory. This presents a strong case for the synthesis of controllers in latent space. For the unconstrained case the MIMO controller design problem is reduced to several SISO design problems. It should be noted here that though the controller design is SISO in the latent space, these univariate controllers are capable of handling the interactions in the original space.

The univariate controllers C_1 through C_n are designed based on the PLS inner models i.e., C_i is designed based on G_i $(i = 1, ..., n)$ using any of the available alternatives (e.g. IMC, pole placement, frequency response techniques). T is the vector of scores computed by the controllers. The scores are then transformed into the real physical inputs which drive the process. Figure 4.3 illustrates the PLS based controller implementation. Thus, the controller *sees* the error signals and the command signals in terms of the basis defined by the columns of the respective loading matrices (Q and P). Because the dynamic part of the PLS model has a diagonal structure, the choice of the input-output pairings is automatic and is optimal in some sense. For further details of PLS based controller implementation the reader is referred to Kaspar and Ray (1992, 1993) and Lakshminarayanan (1996). The latter also discuss constrained control via PLS based dynamic models.

Remark 4.3. *For square as well as non-square systems, pairing in terms of the latent variables is automatic. For* $n < \min(n_u, n_y)$ *there is a dimensionality reduction of the original*

system. For such cases the mapping from the original variables to the latent variables is not exact and a subsequent loss in information results. For controller synthesis in the PLS framework this means that the correlations are taken care of and the controller design and implementation will be in a reduced dimensional latent space.

In this chapter we focus on using the PLS based dynamic models for (i) interactor estimation and multivariable time delay compensation; constrained model predictive control of (ii) linear systems and (iii) a class of nonlinear systems which can be modeled via Hammerstein and Wiener models.

This chapter is organized as follows. Section 4.2 deals with the interactor estimation through PLS based dynamic models. Several desirable properties of the PLS models, from a control point of view, are highlighted. The multivariable time delay compensation strategy based on SISO Smith predictors for each PLS based controller is outlined. The usefulness of the proposed approaches is demonstrated via two simulation examples. A method for constrained model predictive control in the PLS framework is presented in Section 4.2. The constraint mapping from the original to the latent space is highlighted and the proposed approach is demonstrated on a simulation example and a pilot scale stirred tank heater. Extensions to the nonlinear MPC using PLS based Hammerstein and Wiener models are presented in Section 4.3. Application to a pH-level control problem is used to show the efficacy of proposed control technique. Section 4.5 gives the concluding remarks.

With respect to the notation, in Section 1 we used the standard time-series notation, wherein Y, U, V, T were row vectors; in the following Sections we will use standard control literature notation wherein Y, U, V, T are column vectors with appropriate dimensions.

4.2 Interactor Estimation and Multivariate Time Delay Compensation via PLS models

Time delays are common place in chemical processes and often pose problems in controller design and implementation. For SISO systems time delay can be easily separated and its effect can be analyzed. In MIMO systems time delays are difficult to factorize and their interpretation is not as straightforward. The *interactor matrix* is a generalization of the time delay term, for multivariable, discrete systems. In 1976, Wolovich and Falb proved the existence of the interactor matrix in a particular form, i.e., the lower triangular form. The interactor matrix plays an important role in controller design (Sripada, 1988; Walgama, 1986 ; Garcia and Morari, 1982), performance assessment (Harris *et al.*, 1997, Huang *et*

al., 1996), optimal input-output pairing (Tsiligiannis and Svoronos, 1988Tsiligiannis and Svoronos, 1988). A formal definition of the interactor matrix is provided below (Huang, *et al.*, 1996).

Definition 4.1. *For any $n \times n$ proper, rational polynomial transfer function matrix G, there exists a unique, non-singular $n \times n$ lower left triangular polynomial matrix D, such that $det(D) = z^r$ and*

$$\lim_{z^{-1} \to 0} DG = \lim_{z^{-1} \to 0} G_o = K$$

where K is a full rank matrix, the integer r is the number of infinite zeros of G, and G_o is the delay free part of the transfer function matrix $G(z^{-1})$ and contains only the finite zeros of G,

$$G = D^{-1} G_o.$$

The matrix D is defined as the interactor matrix and D^{-1} is called the inverse interactor matrix.

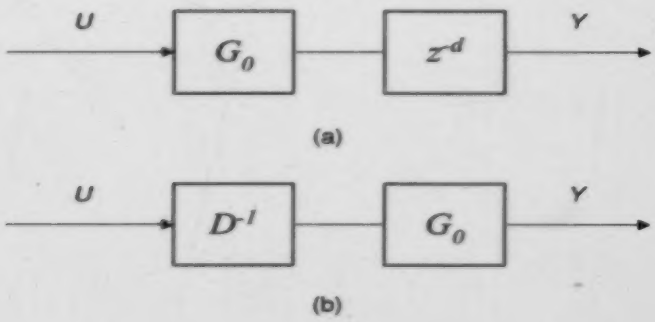


Figure 4.4: *The delay in the (a) SISO case and (b) the multivariable equivalent*

For a given process, the interactor matrix is non-unique and several estimation algorithms have been proposed. The algorithms of Wolovich and Falb (1976) and Goodwin and Sin (1984), require prior knowledge of the transfer function matrix of the process. Shah *et al.* (1987) give a method for estimation of the interactor matrix through a solution of a set of linear algebraic equations of certain Markov parameters. Huang *et al.* (1996), extended this approach to the estimation of an unitary interactor under closed loop conditions. The *best* form of the interactor matrix is application dependent.

Example 4.1. *The first two examples are from Tsiligiannis and Svoronos (1992) and the last two examples are from Huang (1997).*

1) Simple Interactor:

$$G(z) = \begin{bmatrix} 2z^{-d_1} & z^{-d_1} \\ z^{-d_1} & z^{-d_1} \end{bmatrix} = z^{-d_1} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

2) Diagonal Interactor:

$$G(z) = \begin{bmatrix} 2z^{-d_1} & z^{-d_1} \\ z^{-d_2} & z^{-d_2} \end{bmatrix} = \begin{bmatrix} z^{-d_1} & 0 \\ 0 & z^{-d_2} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

3) Triangular Interactor:

$$G(z) = \begin{bmatrix} \frac{z^{-1}}{1+z^{-1}} & \frac{z^{-1}}{1+2z^{-1}} \\ \frac{z^{-1}}{1+3z^{-1}} & \frac{z^{-1}}{1+4z^{-1}} \end{bmatrix}, D = \begin{bmatrix} z & 0 \\ -z^3+2z^2 & z^3 \end{bmatrix}.$$

4) General Interactor:

$$G(z) = \begin{bmatrix} \frac{z^{-1}}{1+z^{-1}} & \frac{z^{-1}}{1+2z^{-1}} \\ \frac{z^{-1}}{1+3z^{-1}} & \frac{z^{-1}}{1+4z^{-1}} \end{bmatrix}, D = 0.5 \begin{bmatrix} z+z^2 & z-z^2 \\ z^2-z^3 & z^2+z^3 \end{bmatrix}.$$

For examples 3) and 4) it is easy to verify that, , where K is full ranked. These two examples also illustrate the non-uniqueness of the interactor. The interactor derived in example 4) is an unitary interactor, $D^T(z^{-1})D(z) = I$. The unitary interactor is known to play a very important role in controller design and multivariate performance assessment (Huang, 1997 and Harris et al., 1996).

PLS based dynamic modeling results in a decoupled system in the latent space. Consequently, the interactor matrix in the latent space is a diagonal matrix and can be easily obtained by factoring out the delays in each channel. A simple transformation yields the interactor matrix in the original space. In terms of the latent variables we have,

$$\begin{aligned} V(z^{-1}) &= \begin{bmatrix} z^{-d_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & z^{-d_n} \end{bmatrix} G_o(z^{-1})T(z^{-1}) \\ &= D^{-1}G_oT. \end{aligned} \tag{4.10}$$

Thus $D^{-1}(z^{-1})$ is the inverse interactor matrix in the latent space and contains all the zeros at infinity. $G_o(z^{-1})$ is the delay free part of the transfer function,

$$G(z^{-1}) = D^{-1}(z^{-1})G_o(z^{-1}).$$

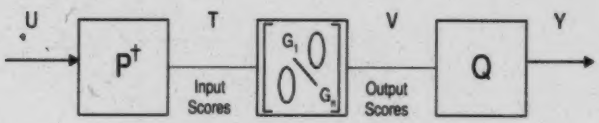


Figure 4.5: *PLS based represenation of dynamic systems*

Now we apply simple linear transformations to express the model in the original basis space,

$$Y(z^{-1}) = QD^{-1}(z^{-1})G_o(z^{-1})P^{\dagger}U(z^{-1}) \tag{4.11}$$

i.e., the original transfer function matrix in the Euclidean basis space is

$$\bar{G} = QD^{-1}G_0P^{\dagger}.$$

Proposition 4.1. *The matrix given by $QD^{-1}(z^{-1})Q^{\dagger}$ represents the inverse interactor matrix in the original domain and will contain all the zeros at the infinity.*

Proof. The poles of $D^{-1}(z^{-1})$ given by

$$\det(D^{-1}(z^{-1})) = z^{-(d_1+d_2+\ldots+d_n)} = 0.$$

Therefore

$$\begin{aligned} \det(QD^{-1}(z^{-1})Q^{\dagger}) &= \det(Q)\det(D^{-1})\det(Q^{\dagger}) \\ &= \det(D^{-1}(z^{-1})). \end{aligned} \tag{4.12}$$

Hence the poles of $D^{-1}(z^{-1})$ are the same as $QD^{-1}(z^{-1})Q^{\dagger}$ and correspond to the infinite zeros of the transfer function matrix.

Remark 4.4. *The interactor matrix thus obtained has a unity gain at steady state.*

Proof. : The inverse interactor matrix in the original domain is defined by

$$\begin{aligned} \bar{D}^{-1}(z^{-1}) &\triangleq QD^{-1}(z^{-1})Q^{\dagger} \\ \therefore \bar{D}^{-1}(1) &= QD^{-1}(1)Q^{\dagger} \\ &= QIQ^{\dagger} \\ &= I. \end{aligned}$$

Remark 4.5. *$G_o(z^{-1})$ can be easily transformed to the Smith McMillan (for definition see Chapter 6 of Kailath, 1980) form of the original transfer function and the poles and zeros of the system, for square systems, are given by,*

$$Zeros: \prod_i n_i(z^{-1}) = 0$$

$$Poles: \prod_i d_i(z^{-1}) = 0$$

where

$$G_o(z^{-1}) = \begin{bmatrix} \frac{n_1(z^{-1})}{d_1(z^{-1})} & 0 & \cdots & 0 \\ 0 & \frac{n_2(z^{-1})}{d_2(z^{-1})} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{n_n(z^{-1})}{d_n(z^{-1})} \end{bmatrix}$$

$$= BA^{-1}. \tag{4.13}$$

The zeros and poles of $G_o(z^{-1})$ are unchanged when a linear transformation is applied. The locations of poles and zeros contained in G_0 are invariant to pre and post-multiplication by square, constant matrices.

Thus the interactor matrix is estimated as a part of the PLS identification process at the cost of an incremental effort. No additional effort, other than pre- and post - multiplication by constant matrices, is required to estimate the interactor matrix. The diagonal nature of the interactor matrix in latent space gives a new interpretation to the interactor matrix in general.

Techniques such as Smith Predictor control scheme (Smith, 1957) and the Dahlin algorithm (Dahlin, 1967) can be used to compensate for time delays in SISO systems. Ogunnaike and Ray (1979) generalized the Smith Predictor approach to multivariable, continuous time systems in such a way that the characteristic closed loop equation is free of the delay terms. Jerome and Ray (1986) extended this method so that the closed loop system had one or more of the following properties:

1. Delay free closed loop characteristic equation
2. The feedback term is a prediction of the control action on the process output
3. The plant is factored into a delay term and a delay free transfer function part.

Property 3 is similar to the interactor factorization discussed in the Section 2. However, their approach is not general and the delay-structure is user defined. Garcia and Morari (1982) propose an inner-outer factorization (T_+T_-) which is similar to the interactor factorization. Their method also separates the non-minimum phase zeros along with the delays

and a multivariable time delay compensator can be designed within the internal model control (IMC) framework . Note that the methods of Ogunnaike and Ray (1979), Jerome and Ray (1986) and Garcia and Morari (1982) can be applied to both, continuous and discrete systems, whereas the interactor factorization is defined only for discrete systems. Figure 4.6 shows the Smith predictor structure for the SISO case, while figure 4.7 illustrates the multivariable delay compensation structure based on the interactor factorization.

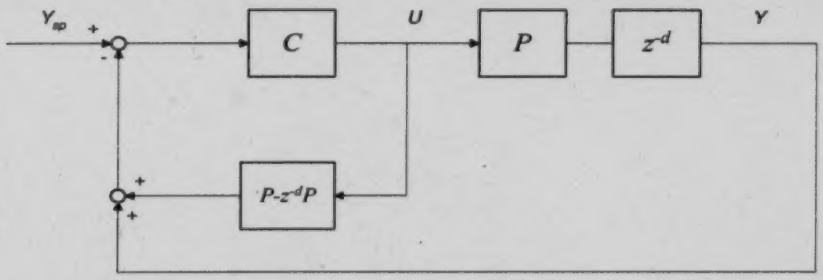


Figure 4.6: *The SISO Smith Predictor structure*

The decoupled structure inherent in PLS models lends itself naturally to a simple dead time compensation strategy based on the classical Smith predictor. We exploit the structure of the PLS model to design n Smith predictors, one for each controller in the latent space. The combined effect of this scheme is a multivariable Smith predictor control scheme. A similar idea was proposed by Kaspar and Ray (1992), but they had the same delays in each of the latent space models. Thus their PLS based Smith predictor is not equivalent to a true multivariate Smith Predictor. In the case of no model plant mismatch (MPM) our dead time compensator will satisfy all of the above three properties. The latent space based Smith Predictor control structure is illustrated in Figure 4.8. $C(z^{-1})$ is diagonal matrix with n SISO controllers, $C_i(z^{-1})$, on the diagonal. Similarly G_o and D^{-1} are diagonal matrices obtained from the PLS based model as per Section 4.1

Proposition 4.2. *It can be shown that PLS based dead time compensation approach is equivalent to the multivariable time delay compensation based on the interactor factorization.*

Proof. The PLS based Smith predictor is given by

$$v_p = (G_o - D^{-1}G_o)t.$$

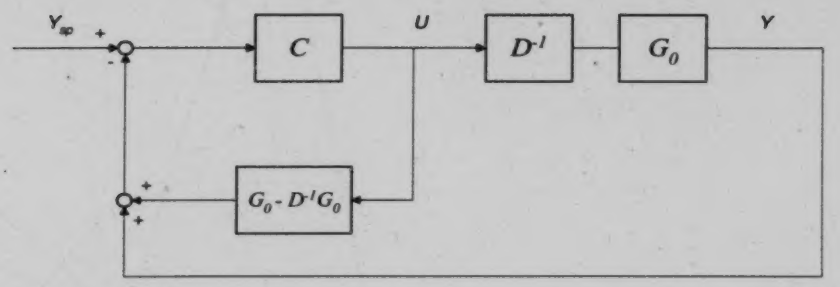


Figure 4.7: *The interactor factorization based multivariable dead time compensation scheme*

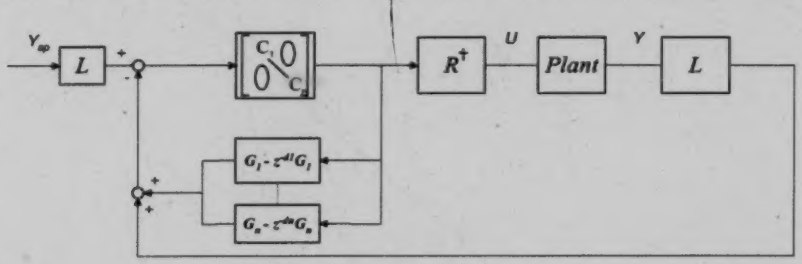


Figure 4.8: *Multivariable delay compensation in the latent space*

Transforming to the original domain we get

$$\begin{aligned} y_p &= L^{-1}(G_o - D^{-1}G_o)Ru \\ &= (\bar{G}_o - \bar{D}^{-1}\bar{G}_o)u \end{aligned}$$

which is the same as the interactor factorization based predictor. Here $\bar{G}_o, \bar{D}^{-1}$ are original domain representations of the delay free transfer function matrix and interactor respectively.For the case of no MPM the PLS based delay compensator satisfies all the three properties of Jerome and Ray (1986).

4.2.1 Numerical Example

The feasibility of the PLS based interactor estimation and dead time compensation strategy is demonstrated on the following numerical example. The example illustrates the interactor estimation via a PLS based dynamic model of a simulated distillation column and shows the usefulness of the multivariable Smith predictor as well.

Example 1: Interactor Estimation and Time Delay Compensation for a Multivariable Process

This methanol-water distillation column model is a 2×2 process with the controlled variables being the distillate (y_1) and bottoms (y_2) compositions of methanol and the manipulated variables being the reflux ratio (u_1) and the reboiler steam flow rate (u_2). The process model was experimentally obtained by Wood and Berry (1973),

$$\begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} = \begin{bmatrix} \frac{12.8e^{-s}}{16.7s+1} & \frac{-18.9e^{-s}}{21s+1} \\ \frac{6.6e^{-7s}}{10.9s+1} & \frac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix}. \tag{4.14}$$

The above model served as the plant in the simulation runs. Identification runs were performed on the simulation model and a discrete PLS based dynamic model was estimated at the sampling rate of 1 minute. This model served as the basis for controller design in latent spaces.

The PLS based dynamic model for the Wood-Berry column is given by

$$V = D^{-1}G_0T$$

where

$$D^{-1} = \begin{bmatrix} z^{-5} & 0 \\ 0 & z^{-5} \end{bmatrix}$$

$$G_0 = \begin{bmatrix} \frac{0.142}{1-0.43z^{-1}-0.47z^{-2}} & 0 \\ 0 & \frac{0.053+0.029z^{-1}}{1-0.23z^{-1}-0.23z^{-2}} \end{bmatrix}.$$

The transformation matrices, including the scaling matrices, are given by

$$P^\dagger = \begin{bmatrix} 0.154 & 0.603 \\ -0.452 & 0.207 \end{bmatrix}, Q = \begin{bmatrix} 0.049 & 0.063 \\ 0.055 & -0.055 \end{bmatrix}.$$

Interactor Estimation: The interactor matrix in the original domain is given by

$$QD^{-1}(z^{-1})Q^\dagger = z^{-5}I.$$

The interactor free part in the original domain is given by $QG_o(z^{-1})P^\dagger$. Let us consider the case where the delay in the first channel was 4 units instead of 5. The interactor matrix now becomes

$$QD^{-1}Q^\dagger = \begin{bmatrix} 0.049 & 0.063 \\ 0.055 & -0.055 \end{bmatrix} \begin{bmatrix} z^{-4} & 0 \\ 0 & z^{-5} \end{bmatrix} \begin{bmatrix} 8.89 & 10.15 \\ 8.81 & -7.99 \end{bmatrix}$$

$$= \begin{bmatrix} 0.443z^{-4} + 0.557z^{-5} & 0.506z^{-4} - 0.506z^{-5} \\ 0.488z^{-4} - 0.488z^{-5} & 0.557z^{-4} + 0.443z^{-5} \end{bmatrix}.$$

Verification:

i) The determinant of the interactor matrix:

$$\begin{aligned} & \det(QD^{-1}Q^{\dagger}) \\ = \; & 0.247z^{-8} + 0.507z^{-9} + 0.247z^{-10} - (0.247z^{-8} + 0.493z^{-9} + 0.247z^{-10}) \\ = \; & z^{-9}. \end{aligned}$$

ii) The unity DC gain property:

$$\begin{aligned} & \bar{D}^{-1}(1) \\ = \; & \begin{bmatrix} 0.443 + 0.557 & 0.506 - 0.506 \\ 0.488 - 0.488 & 0.557 + 0.443 \end{bmatrix} \\ = \; & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Multivariable Smith Predictor: The results for the Wood-Berry Column are presented below (see Figures 4.9 and 4.10). In both cases, two PI controllers were designed based on the PLS models,

$$C_i(z^{-1}) = K_i(1 + \frac{T}{\tau_i}\frac{1}{z-1}); i = 1, 2$$

where T is the sampling time. When used with the Smith predictors, higher gain controllers could be used in both cases. The controller tuning parameters were

$$\begin{aligned} \text{With Smith predictor} \quad &: \quad K_1 = K_2 = 5, (T/\tau_I)_i = 1 \\ \text{Without Smith predictor} \quad &: \quad K_1 = K_2 = 2, (T/\tau_I)_i = 1 \end{aligned}$$

For the second run the delay in the plant was reduced by one sampling interval, thus introducing a delay mismatch. The same controllers as above were implemented on the plant and the results are shown in Figure 4.10. The Smith predictor based controller, with its high gains still outperforms the conventional controller with more conservative gains. Similar observations are made by Tan *et al.* (1996), who found that a mismatched Smith-system yields a better performance over a conventional controller without delay compensation.

4.2.2 Discussion

Modeling in latent spaces offers several advantages for controller design purposes over conventional identification methods

- The dynamics are decoupled leading to a diagonal transfer function matrix. Thus if one wants to design the controller in the original space, the interactor matrix can be obtained via a simple transformation

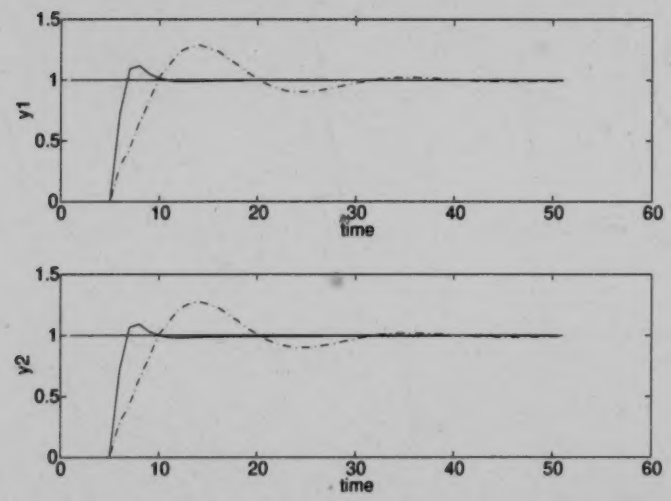


Figure 4.9: *PLS based Smith predictor for the case of no model plant mismatch (– with SP, - - without SP)*

- Non-minimum phase zeros and unstable poles can also be factored out along with the delays in a transparent manner.
- The interactor matrix thus obtained is a unity gain matrix.
- The poles and zeros of the system can be directly found from the PLS model.
- A left coprime factorization of the transfer function is also obtained.
- A SISO Smith predictor scheme can be employed in the latent variable subspace to do multivariable time delay compensation. It was shown that this approach is equivalent to the interactor factorization based dead time compensation in the original domain.

These attributes of the PLS based modeling and control scheme were demonstrated through two simulation examples. The robustness of the Smith predictor based dead time compensation was emphasized. Thus the PLS based model has properties close to that of a *canonical form* and is unique in several ways. No single identification scheme has so many desirable features from a control point of view. The accuracy of the PLS estimates, however, is not easy to quantify (Phatak *et al.*, 1993). Quantification of the uncertainty in the PLS estimates could be utilized in robust PLS based controller synthesis. In the next Section, constrained model predictive control of linear systems using PLS based dynamic models is presented.

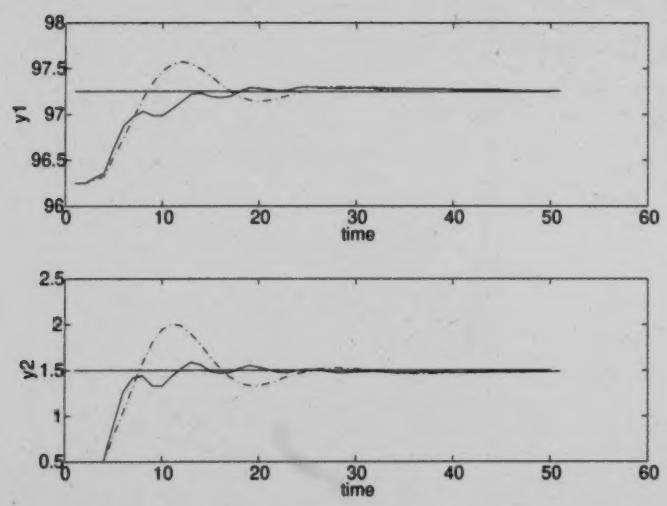


Figure 4.10: *PLS based Smith predictor for the case of a delay mismatch (– with SP, - - without SP)*

4.3 Model Predictive Control in Latent Spaces

Model Predictive Control (MPC) schemes are gaining increasingly wide acceptance in the chemical process industries. Various forms of model predictive control schemes have been reported in the last decade owing to great interest evinced both by industrial practitioners and academic researchers. Besides their ability to handle large multivariable systems, the key feature that makes these algorithms attractive for industrial applications is their ability to handle constraints that arise in any real life process control situation. To achieve higher profits, the supervisory control layer often forces the process to operate at the intersection of constraints. MPC schemes incorporate the constraints explicitly in the control design strategy by solving a constrained optimization problem at each sampling instant and implementing the control moves in a receding horizon fashion. A variety of process descriptions ranging from step/impulse response coefficients, discrete transfer function models to state space models are employed by these MPC schemes - each having its merits and drawbacks.

A complete process model must be available in order to implement any of these control schemes. The multivariable process model is usually obtained empirically through an identification experiment. In the identification of MIMO processes, a high degree of correlation is often observed between process variables. In such cases, use of identification software based on the ordinary least squares (OLS) technique will result in parameter estimates with

large variances owing to the ill-conditioned nature of the problem. Multivariate statistical techniques such as PLS have proved to be robust when presented with such correlated data. Consequently PLS and related methods (e.g. Principal Components Analysis (PCA), Ridge Regression (RR)) have found applications in the area of process monitoring, fault detection and identification (Kresta (1992), Wise (1991), Qin and McAvoy (1992), Kaspar and Ray (1992,1993) , Nomikos and MacGregor (1995)).

In this Section, we combine the PLS based multivariable modelling strategy proposed by Lakshminarayanan *et al.* (1996) with the MPC algorithm (Cutler and Ramaker, 1980)? in order to provide constrained control of multivariable systems. Different ways of dealing with process constraints will be explored - their merits and drawbacks will be pointed out. Illustrative examples involving the simulation of the Wood-Berry column and an experimental laboratory stirred tank heater will be presented.

4.3.1 PLS based constrained MPC of Linear Systems

The MPC objective function objective function in terms of the original variables is given by

$$\begin{aligned} \min J &= \Delta u_k^T (S^T \Gamma S + \Lambda) \Delta u_k - 2(r_k - f_k - d_k)^T \Gamma S \Delta u_k \\ &\text{subject to } A\Delta u_k + B \geq 0 \end{aligned} \tag{4.15}$$

where S is the dynamic matrix comprising the step response coefficients, f_k is the free response vector, d_k is the estimate of the disturbance vector and r_k is the setpoint trajectory that we want to achieve. For a more detailed account of the terms involved in the MPC formulation, see Chapter 2. In the latent space the same problem can be reformulated in terms of the latent inputs and outputs. In this case the MPC objective function is based on the linear dynamic model between t (inputs) and v (outputs)

$$\begin{aligned} \min J_L &= \Delta t_k^T (S_L^T \Gamma' S_L + \Lambda') \Delta t_k - 2(r_k - f_k - d_k)^T \Gamma S_L \Delta t_k \\ &\text{subject to } A'\Delta t_k + B'_k \geq 0 \end{aligned} \tag{4.16}$$

where the primed quantities are the corresponding expressions in terms of the latent variables and S_L is the dynamic matrix obtained from the PLS dynamic model in the latent space. The free response and the disturbance estimate are also obtained from the PLS model. The constraints are posed in the original space and transformed as per the previous Section.

Note that the latent variables are scaled variables, hence the control and output weightings have to be chosen accordingly. The time scales however, are invariant to the transformations and therefore the choice of the prediction and control horizons - p & m - will remain the same in both the original and the latent spaces. In this Section we will discuss the constraint mapping when we pose the problem in terms of the latent variables i.e. how to arrive at A',B'. The individual transformations relating the original to the latent variables are given by

$$\begin{aligned} t &= P^{\dagger}u \\ y &= Qv \end{aligned} \tag{4.17}$$

where $P^{\dagger}$ is the generalized inverse of P.

Consider a 2×2 system which is modeled using 2 PLS dimensions. The constraints in the original variables are decoupled if we restrict ourselves to input and slew rate constraints, e.g. constraints on u_1 and u_2 are independent of each other. For the sake of brevity only the two variable case is illustrated here, but the concepts presented can be quite easily extended to the higher dimensional case. To illustrate our point geometrically let us look at the region where $u_1$and u_2 are required to remain.
This region can be represented by the inequality

$$\begin{aligned} u_{\min} &\leq u \leq u_{\max} \Leftrightarrow \\ u_{\min} &\leq Pt \leq u_{\max}. \end{aligned} \tag{4.18}$$

Expanding we get

$$\begin{bmatrix} u_1^{\min} \\ u_2^{\min} \end{bmatrix} \leq \begin{bmatrix} p_{11}t_1 + p_{12}t_2 \\ p_{21}t_1 + p_{22}t_2 \end{bmatrix} \leq \begin{bmatrix} u_1^{\max} \\ u_2^{\max} \end{bmatrix}. \tag{4.19}$$

This region is shown in figure 4.12. Thus the constraints get *coupled* in the latent space. The rate (Δu) constraints get affected similarly. The main outcome of posing the constraints in the latent space is that the problem gets coupled through the constraints in spite of the dynamics being decoupled in the latent space. Therefore a multivariate approach to controller synthesis is needed to handle the constraints. In this situation SISO theory cannot be used to design controllers in the latent space. If the output constraints are being used then the constraints are coupled irrespective of the transformations. Often, the output constraints are excluded since they often lead to infeasibility problems in the optimization step.

In the case of linear systems, equation (4.11) provides a model in terms of the original variables. For the unconstrained case, the dynamic inner models can be used in two ways: (1) Each inner model can be used to develop SISO MPC controllers and (2) A MIMO MPC controller which utilizes all the n inner models together. If constraints are imposed on the manipulated variables, the constraints in the latent space are coupled (as described below). If strategy (1) is employed, then the controllers must act in a co-ordinated fashion. Otherwise, constraints on the original variables will be violated. With strategy (2), a one-time transformation of the constraints is adequate for efficient implementation of the MPC algorithm.

In the original space, the constraints are represented as:

$$u_{min} \leq u(k) \leq u_{max} \tag{4.20}$$

$$\Delta u_{min} \leq \Delta u(k) \leq \Delta u_{max} \tag{4.21}$$

$$y_{min} \leq \hat{y}(k) \leq y_{max}. \tag{4.22}$$

For a case involving two manipulated variables, the amplitude constraints are shown in figure 4.12(a) (mathematically expressed in equation (4.20)). In terms of the latent space variables and the PLS matrices, the above equation may be written as:

$$u_{min} \leq Pt(k) \leq u_{max}. \tag{4.23}$$

A graphical plot of the constraints in terms of the input space latent variables (T) is depicted in figure 4.12(b). Use of the PLS inner models with the constraints as given in equation (4.23), will ensure the satisfaction of constraints in the original space. Such a mapping is *one to one* - each point in the constrained original space has a unique image in the constrained latent space and vice versa. The outcome of transforming the original constraints into latent space constraints is that the constraints that were decoupled in the original space become coupled in the latent space. A similar analysis holds for the rate constraints as well. Hence a multivariate approach to controller design is mandatory.

Let us pose the problem of mapping constraints differently. The constraints are now posed in the latent space in a decoupled form. We now seek the maximum and minimum values in the t-space, t_{max} and t_{min}, such that the constraints in the original space are satisfied i.e., find $t_{min} \leq t(k) \leq t_{max}$ such that $u_{min} \leq u(k) \leq u_{max}$. This approach results in original space constraints that are coupled. When the constrained regions of the previous approach and this approach are plotted together, we notice the suboptimality of this approach (see figure 4.13) - not all of the original constraint region is utilized (space

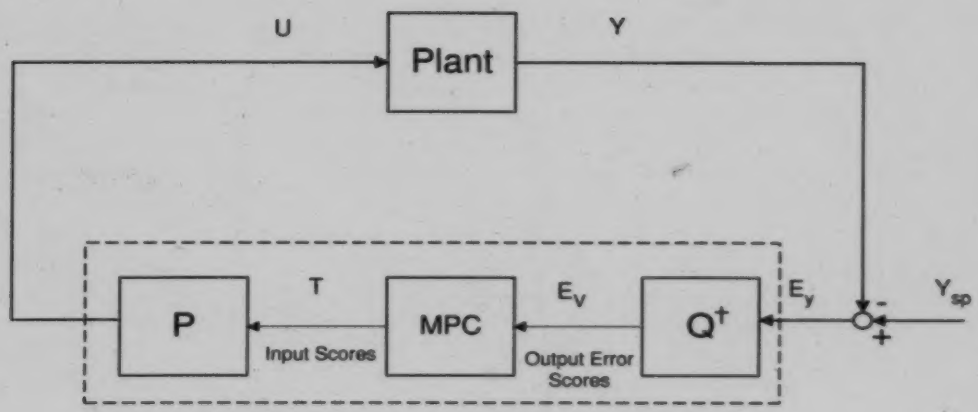


Figure 4.11: *Linear MPC based on Dynamic PLS models*

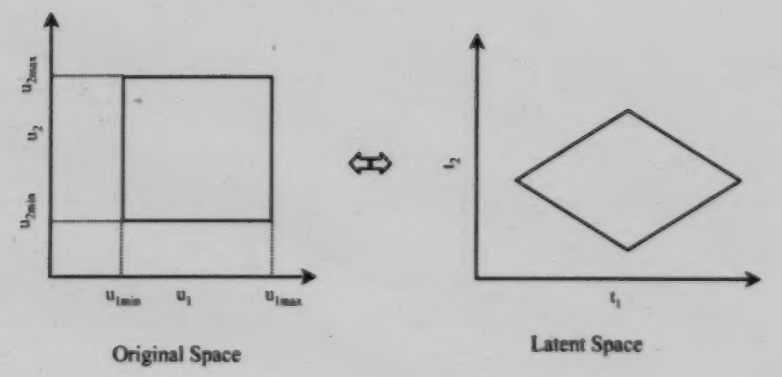


Figure 4.12: *Constraint regions in the original and latent space.*

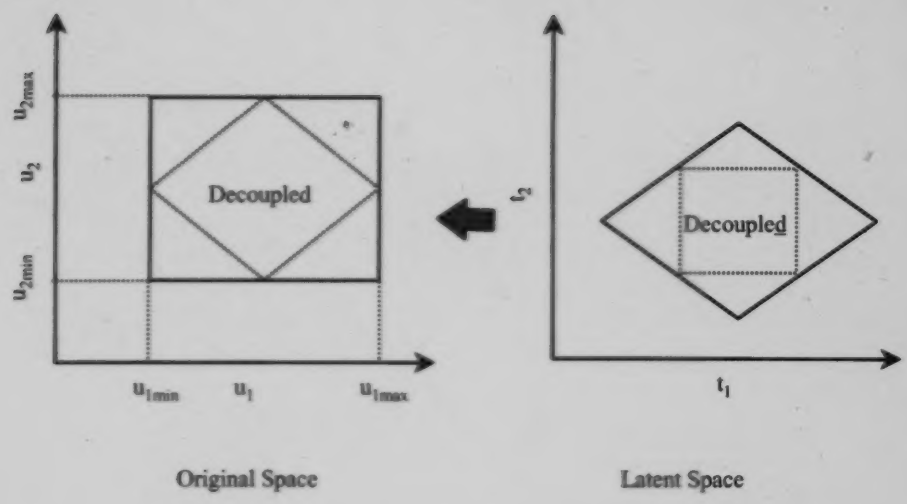


Figure 4.13: *Transformations of the constraint region from the original space to latent input space - necessary and sufficient conditions.*

bounded by broken lines) because we seek to match only the necessary conditions i.e. the maximum and minimum values. The constraints are satisfied but the controller does not use some *permitted* regions in the input space implying that some set points cannot be reached and some disturbances will not be rejected completely.

Simulation of the Wood-Berry Column

We revisit the distillation column model of Wood and Berry (1973) considered in Section 4.2. The model in equation 4.14 serves as the *plant* in the simulation runs. For this system, Lakshminarayanan *et al.* (1996) developed a dynamic PLS model using input-output data generated from equation (4.14). The PLS model is,

$$P = \begin{bmatrix} 0.3228 & 0.9455 \\ -0.9465 & 0.3256 \end{bmatrix}, Q = \begin{bmatrix} 0.3256 & 0.9465 \\ -0.9455 & 0.3228 \end{bmatrix}$$

$$G_1 = \frac{0.1417z^{-5}}{1 - 0.4305z^{-1} - 0.4706z^{-2}}$$

$$G_2 = \frac{0.0529z^{-5} + 0.0291z^{-6}}{1 - 0.2336z^{-1} - 0.2321z^{-2}}.$$

The comparison of the conventional and the PLS based MPC (figure 4.14) clearly indicates the effect of scaling on the control weighting Λ. $\Lambda = 100I$ in the original space yields the same performance as $\Lambda' = 0.05I$ in the latent space (ISE values were compared for servo as well as several regulatory runs). Since the variables are well scaled in the latent space and the dynamics are decoupled, the choice of tuning parameters - Γ, Λ is easier.

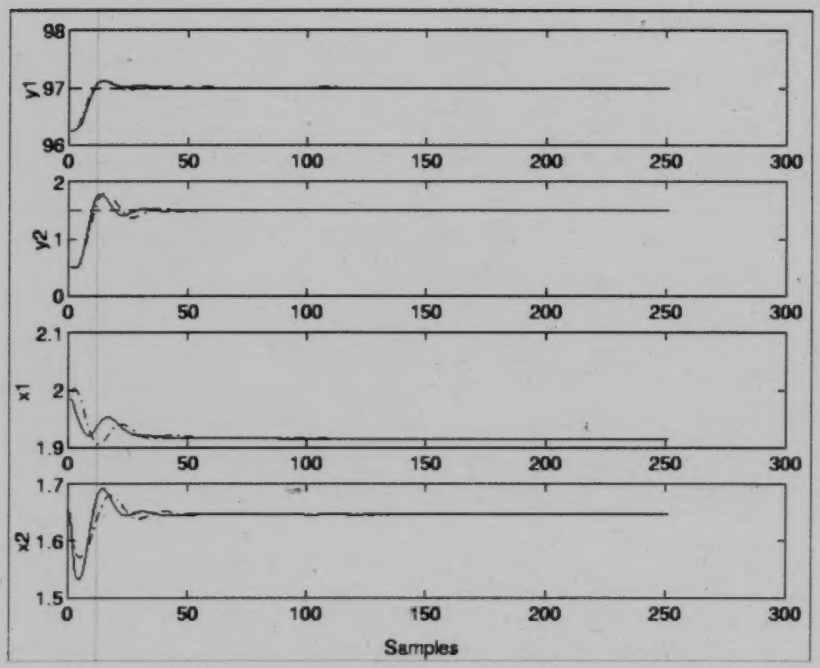


Figure 4.14: *Effect of themove suppression factor on the tracking performance of the PLS based MPC (solid line) and conventional MPC (dashdotted line).*

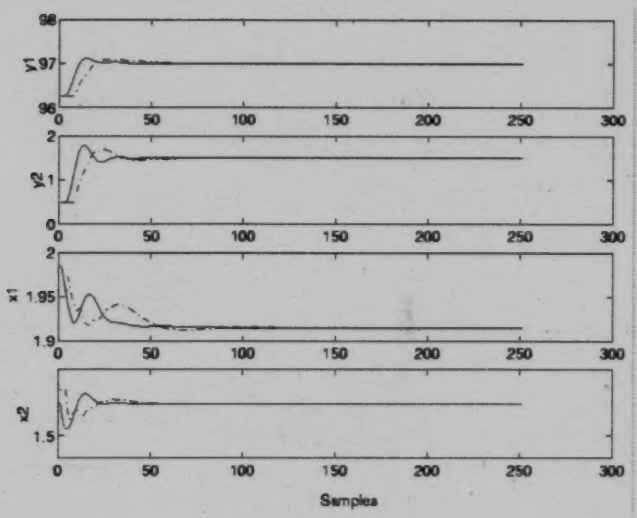


Figure 4.15: *Comparison of the PLS based MPC (solid line) with PID control (dash dotted line).*

In figure 4.15, we compare the performance of the MPC to that of the Vogel-Edgar digital controller. Both the controllers were implemented in the latent space. The results indicate the superior performance of the MPC controller.

Real-Time Control of the Laboratory Stirred Tank Heater

The laboratory stirred tank heater is a cylindrical tank of uniform cross Section where two streams of water (one hot and the other cold) are mixed. The contents of the tank exit through a long and winding copper tube. The flow rates of the hot and cold water streams serve as manipulated variables to control the level of water in the tank and the temperature of the exit stream. Facilities exist to introduce disturbances in the steam flow through the heater coils, the inlet temperature of the hot water stream *etc.* The control algorithm was implemented using a personal computer running Real-Time Matlab/Simulink.

The results of the laboratory run showing servo and regulatory responses is depicted in figure 4.16. Two set point changes in each output variable were made. An unknown amount of cold water was dumped into the vessel around the 1575th sample point. Satisfactory servo and regulatory responses were obtained by employing $m = 2$, $p = 20$ and $\Lambda' = 10I$ as the controller parameters. The constraints placed on the amplitude (0-100%) and rate ($\pm$10%) of the manipulated variables remained inviolate over the entire duration of the experiment.

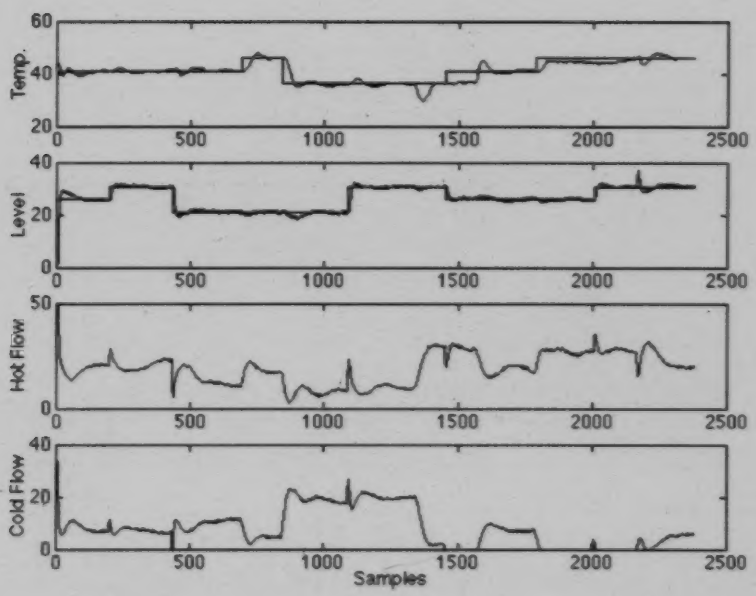


Figure 4.16: *Experimental evaluation of the PLS based MPC scheme on a pilot scale stirred tank heater.*

4.4 Constrained Nonlinear Model Predictive Control using PLS based Hammerstein and Wiener Models

Most chemical processes are inherently nonlinear. Researchers have proposed several ways to equip MPC with the capability to deal with nonlinear processes. The approaches to nonlinear MPC (NLMPC) can be categorized into two groups: (1) those based on first principles models of the process and (2) those based on black box models identified from input-output data.

If the MPC objective function is to be minimized using the first principles, nonlinear model, a dynamic optimization problem has to be solved at every sampling instant. Biegler (1984) proposed approximation of the differential equations using orthogonal collocation in order to reduce the optimal control problem to a nonlinear program which could be solved using sequential quadratic programming (SQP). Patwardhan *et al.* (1990) extended Biegler's approach to the constrained nonlinear MPC problem. This technique requires selection of appropriate orders for the polynomial approximations and increases the size of the optimization problem manifold. Economou and Morari (1986) adopted the nonlinear internal model control (IMC) framework wherein the nonlinear model inverse was obtained by a Newton type method. Li and Biegler (1988,1989) improved the convergence properties of this method by including a relaxation parameter. Constraints were incorporated through a quadratic programming (QP) formulation.

Brengel and Seider (1989) proposed a NLMPC scheme using successive linearization of the model equations at each sampling interval. This involves analytical/numerical evaluation of the Jacobians and discretization. The linearized model is used for predictions and only a quadratic program has to be solved at each iteration. Garcia (1984), used the full nonlinear model for finding the effect of the past inputs and estimated disturbances on the output and the linearized model for predicting the effect of the future inputs. Thus the objective function remains quadratic and the computational burden is minimal. Gattu and Zafirou (1992) extended this method by incorporating state estimation using a Kalman filter for better disturbance rejection. They successfully applied this scheme to a semi-batch polymerization reactor and an unstable open loop process.

Patwardhan and Madhavan (1993) approximated the model equations using a second order Taylor series expansion around the steady state. For processes showing highly nonlinear behavior (e.g. input multiplicities) the second order approximation proves superior in comparison to a linear model that includes only first order terms. Mutha *et al.* (1997) pro-

posed a nonlinear MPC scheme for control non-affine systems based on a re-interpretation of the prediction equation as a Taylor series expansion.

In a novel strategy based on the first principles model, Henson and Kurtz (1994), Oliveira, *et al.* (1995) proposed feedback linearization of the nonlinear process followed by linear MPC of the feedback linearized loop. Though the model used in MPC design is linear, the nature of the mappings results in the constraints being nonlinear. Oliveira *et al.* (1995) propose a suboptimal strategy in order to preserve the linearity of the constraints. Henson and Kurtz (1994) formulate the constraint mapping as an optimization step, but sacrifice optimality to maintain linearity of the constraints.

Fundamental models of large chemical processes involve a significant effort and rely on complete knowledge of the process, plant operating conditions and parameters. This kind of knowledge is often scarce and a practical alternative is the use of nonlinear identification procedures to obtain the nonlinear mapping from input-output data.

Bhat and McAvoy (1990) developed artificial neural network (ANN) models of the nonlinear plant under consideration and proposed use of the ANN models for prediction in MPC. Though ANN models are known to approximate any type of nonlinear behaviour, the modeling effort is large and a difficult nonlinear program has to be solved at every instant in order to calculate the control moves.

Hernandez and Arkun (1993) used polynomial ARMA models to capture the nonlinearities; the model considers cross terms between past inputs and outputs and the estimation problem is posed as a linear least squares problem. The MPC objective function is nonlinear with the polynomial ARMA model as the predictor. Doyle *et al.* (1995) and Manner *et al.* (1996) use second order Volterra models to describe nonlinear behavior. The Volterra model considers the cross terms between the past inputs in the manner of convolution models. A large number of coefficients are required for modeling purposes and the optimization step is a nonlinear program. Carlemann-Rugh linearization of the mechanistic process model enables estimation of the model parameters without resorting to experimentation.

Hammerstein and Wiener models are block oriented nonlinear models in which a linear dynamic element is cascaded in series to a static nonlinear element. For Hammerstein models, the nonlinear block precedes the linear dynamic element; for Wiener models the order of the linear and the nonlinear blocks is reversed. Zhu and Seborg (1994) use a modified Hammerstein model to do unconstrained MPC of a univariate pH control process. Norquay *et al.* (1996,1997) demonstrate the use of a Wiener model based MPC on the same pH process. Norquay *et al.* (1997) propose a heuristic scheme for handling constraints.

In this work a novel approach to input constrained nonlinear MPC using partial least squares (PLS) based Hammerstein and Wiener models is presented. It combines a powerful statistical tool - PLS and two elegant representations of nonlinear systems namely Hammerstein and Wiener models with an advanced multivariate controller - MPC. Conventional MIMO Hammerstein/Wiener models present numerical difficulties for multivariable control. For control purposes, solution of a set of nonlinear algebraic equations is involved and this can lead to numerical instability. The use of PLS framework decomposes the modeling problem into a series of univariate problems in the *latent subspace*. The controller synthesis is carried out in the *latent subspace* and facilitates the use of a series of SISO Hammerstein and Wiener models, for MIMO processes.

The main contribution of this Section is to show that the proposed approach is effective in handling multivariable processes and does not compromise the optimality of the constraints. It is shown that, for Wiener models the input constraints preserve their linearity in the latent space and only a quadratic program has to be solved to find the optimal input moves. For the Hammerstein model the input constraints get transformed into a nonlinear region in terms of the latent variables and a nonlinear program results. Henson and Kurtz (1994), Oliveira *et al.* (1995) faced a similar situation with their feedback linearization+MPC scheme but they opted for a suboptimal feasible approach to keep the constraint set linear. The proposed constrained multivariable NLMPC strategies are demonstrated on a simulation case study involving the pH-level control of a nonlinear acid-base neutralization process.

4.4.1 Nonlinear Modeling using PLS based Hammerstein and Wiener models

For nonlinear systems a natural extension of the dynamic PLS approach proposed by Lakshminarayanan (1997a) leads to the use of nonlinear models for explaining the inner relationship between the scores. In general any nonlinear modeling approach like ANN, polynomial ARMA can be used. In this Section we discuss Hammerstein and Wiener structures for modeling the inner relationship. The two approaches are best illustrated by figures 4.18 and 4.17.

The Hammerstein model structure consists of a nonlinear algebraic element followed by a linear dynamic element. In contrast, the Wiener model configures the nonlinear static element following the linear dynamic element. See Haber and Unbehauen (1990) for a comparison of Hammerstein and Wiener structures with other input-output approaches to nonlinear identification. Eskinat *et al.* (1991) used the Hammerstein structure for modeling

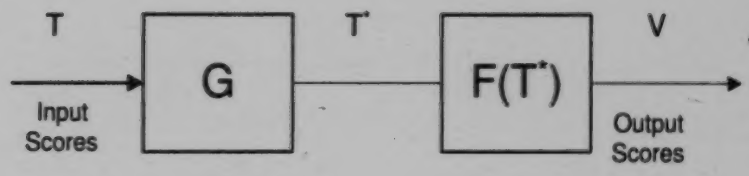


Figure 4.17: *Wiener Models in Latent Spaces*

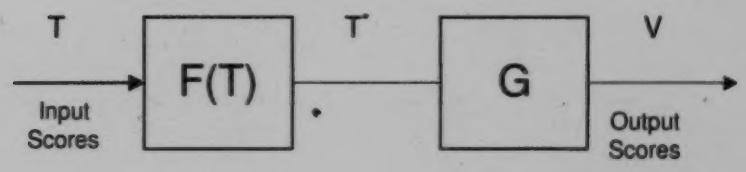


Figure 4.18: *Hammerstein Modeling in Latent Spaces*

a simulated distillation column and an experimental heat exchanger. Lakshminarayanan *et al.* (1995) extended the cannonical variate analysis technique, a subspace identification method, to the estimation of MIMO Hammerstein models. Verhagen and Westwick (1996) modified the MOESP family of subspace methods for identification of Hammerstein type systems.

PLS based Hammerstein models are obtained by relating the scores at each dimension using a SISO Hammerstein model. A polynomial function is used to capture the nonlinearity of the process. The order of the chosen polynomial is an indication of the degree of nonlinearity of the process. For an elaborate discussion of the PLS based Hammerstein modeling approach the reader is referred to Lakshminarayanan *et al.* (1997a) .

The inner relation between the input and output scores is given by

$$\begin{aligned} \text{Hammerstein Model} \quad &: \quad t_i^* = f_i(t_i); v_i(z^{-1}) = G_i(z^{-1})t_i^*(z^{-1}); i = 1, ..., n \\ \text{Wiener Model} \quad &: \quad t_i^*(z^{-1}) = G_i(z^{-1})t_i(z^{-1}); v_i = f_i(t_i^*); i = 1, ..., n \end{aligned} \tag{4.24}$$

4.4.2 Constrained NLMPC via PLS based Hammerstein and Wiener Models

In this Section we discuss the nonlinear model predictive control using PLS based Hammerstein and Wiener models. The inherent structure present in Hammerstein and Weiner models leads to a very intuitive control strategy. In both cases, a linear MPC is designed with the linear dynamic element serving as the model and the nonlinear operator is inverted.

The order in which the linear MPC and the inversion operation is carried out depends on the model structure (see Figures 4.19 and 4.20). Assuming that the nonlinearity is well captured in the region of operation, this kind of control strategy results in an approximately *feedback linearized* system.

Remark 4.6. *It should be noted here that the inversion of nonlinear operator F is fairly straightforward in the latent space since it is a decoupled system,*

$$F(T) = \left[\begin{array}{cccc} f_1(t_1) & f_2(t_2) & .. & f_n(t_n) \end{array} \right].$$

Each f_i is a polynomial equation in a single variable t_i. Hence inversion is guaranteed since robust numerical techniques are available for solving polynomial equations in one variable. In general the polynomial will have multiple roots, we choose the smallest real root. In this regard it is safe to choose a polynomial with an odd degree to ensure the presence of at least one real root.

Remark 4.7. *In the ideal case of no model plant mismatch, the resulting feedback system is linear and linear stability analysis can be used for assessing the closed loop stability of the nonlinear system. However, in practice, model plant mismatch is inevitable and nonlinear stability analysis will have to be carried out for the general case.*

In the remaining discussion we shall concern ourselves with the application of constrained MPC on nonlinear systems modeled using Hammerstein and Wiener models.

Constraint Mapping

Hammerstein Models

The MPC objective function is based on the linear dynamic element between t^*(inputs) and v (outputs),

$$\min_{\Delta t_k^*} J_H = (\Delta t_k^*)^T (S_L^T \Gamma S_L + \Lambda) \Delta t_k^* - 2(r_k' - f_k' - d_k')^T \Gamma S_L \Delta t^* \quad (4.25)$$

$$\text{subject to } g(t_k^*) \geq 0. \quad (4.26)$$

For the sake of brevity, we use t to denote the $n \times 1$ dimensional latent space vector.

The nature of the transformations involved leads to the transition of the linear constraints in the original space to nonlinear constraints in the latent space. The following steps illustrate our point,

$$A\Delta u_k + B_k \geq 0 \Leftrightarrow$$

$$A'\Delta t_k + B_k \geq 0 \Rightarrow$$

$$A' \begin{bmatrix} F^{-1}(t^*(k)) - F^{-1}(t^*(k-1)) \\ F^{-1}(t^*(k+1)) - F^{-1}(t^*(k)) \\ \vdots \\ F^{-1}(t^*(k+m-1)) - F^{-1}(t^*(k+m-2)) \end{bmatrix} + B'_k \geq 0 \Leftrightarrow$$

$$g(t_k^*) \geq 0. \qquad (4.27)$$

where $g(t_k^*)$ is nonlinear operator. The transformation from u to t is linear and is given by:

$$\begin{aligned} u(k) &= Pt(k) \\ A' &= A\left[diag\left(\begin{array}{cccc} P & P & \ldots & P \end{array}\right)\right]_{n_u m \times n_y m}. \end{aligned}$$

The transformation from t^* to $t : t_i^* = f_i(t_i)$, however, is nonlinear. The inverse operator is not one-to-one and hence the constraint mapping is non-unique.

From equation 4.25, it is clear that computation of the control moves involves solution of a nonlinear program at every instant. The nonlinear program comprises a quadratic objective function and a nonlinear constraint set. The nonlinearity of the constraints arises due to the particular structure of the Hammerstein model.

The steps involved in the implementation of the Hammerstein based NLMPC scheme (see figure 4.19) are enumerated below:

1. At sampling instant k, measure the output $y_m(k)$ and the setpoint $r(k)$.

2. Transform the output and setpoint in terms of latent variables using the following relations: $v(k) = (S_y Q)^{-1} y(k); v_{sp}(k) = (S_y Q)^{-1} r(k)$. Send the transformed variables to the PLS based MPC.

3. The controller is based on the linear dynamic part and will calculate inputs in terms of $t^*(k)$. This is done by solving the nonlinear program using a suitable optimization tool. The initial guesses for the solver are provided by equating the initial values to the previous solution - $t_i^* = t^*(k-1)$. For $k = 1$, the initial guess is set to zero.

4. The inverse operator is used to map these values onto the corresponding $t(k)$: $t(k) = F^{-1}(t^*(k))$.

5. The inputs in terms of the latent variables are mapped onto the inputs in terms of original variables: $u(k) = S_u Pt(k)$.

6. Implement the calculated inputs, $u(k)$, on the plant

Wiener Models

The nonlinearity of the constraints can be avoided if a Wiener model structure is chosen. The Wiener model structure succeeds in preserving the linearity of the constraints in spite of the nonlinear mappings involved. However, this is not true if output constraints are also involved. For the present purposes we restrict ourselves to rate and amplitude constraints only since output constraints often result in infeasibility problems. In this case the MPC objective function is based on the linear dynamic model between t (inputs) and t^* (outputs),

$$\begin{aligned} \min_{\Delta t} J_W &= (\Delta t_k)^T (S_L^T \Gamma S_L + \Lambda)\Delta t_k - 2(r_k' - f_k' - d_k')^T \Gamma S_L \Delta t_k \\ \text{subject to } A'\Delta t_k + B_k &\geq 0. \end{aligned} \tag{4.28}$$

The linearity of the constraints is preserved since the nonlinearity is now on the output instead of the input and the transformation from u to t is a linear one:

$$\begin{aligned} A\Delta u_k + B_k &\geq 0 \Leftrightarrow \\ A'\Delta t_k + B_k &\geq 0. \end{aligned} \tag{4.29}$$

The relationship between A and A' remains the same as before. The geometry of this constraint set is the same as in figure 4.12.

Thus a simple change in the model structure results in high gains in terms of the computational complexity of the optimization problem. Only a quadratic program has to be solved at every instant. In cases where the Wiener modeling approach is applicable, it offers distinct advantages for design of a constrained controller over other NLMPC methods based on black-box models.

Remark 4.8. *In the latent space, Wiener models preserve the linearity of the input constraints. Hammerstein models result in the original linear constraints being mapped to a nonlinear constraint region.*

The steps involved in the implementation of the Wiener based NLMPC scheme (see figure 4.20) are enumerated below

1. At sampling instant k, measure the output $y(k)$ and the setpoint $r(k)$.
2. Transform the output and setpoint in terms of latent variables using the following relations: $v(k) = (S_y Q)^{-1} y(k); v_{sp}(k) = (S_y Q)^{-1} r(k)$

3. Invert the nonlinear operator to express these values in terms of t^* : $t^*(k) = F^{-1}(v(k)); t^*_{sp}(k) = F^{-1}(v_{sp}(k))$. Send the transformed variables to the PLS based MPC.

4. The controller is based on the linear dynamic part and calculates the inputs in terms of $t(k)$. This is done by solving the quadratic program using any standard QP solver.

5. The inputs in terms of the latent variables are mapped onto the inputs in terms of original variables: $u(k) = S_u P t(k)$.

6. Implement the calculated inputs, $u(k)$, on the plant

4.4.3 Discussion

The ability to handle output constraints is one of the main advantages of constrained MPC. The output constraints can be posed in the following manner for Hammerstein and Wiener models respectively:

Hammerstein models:

$$\begin{aligned} \text{Original Variables} &: \quad y_{\min} \leq \hat{y}_k \leq y_{\max} \\ \text{Latent Variables} &: \quad y_{\min} \leq Q^{\dagger} v_k \leq y_{\max} \\ &: \quad y_{\min} \leq Q^{\dagger}\{S_L \Delta t^*_k + f'_k + d'_k\} \leq y_{\max} \end{aligned}$$

where S, f and d are obtained from the linear dynamic element between v and t^*. The output constraints are linear in terms of the latent variables for PLS based Hammerstein models.

Wiener models:

$$\begin{aligned} \text{Original Variables} &: \quad y_{\min} \leq \hat{y}_k \leq y_{\max} \\ \text{Latent Variables} &: \quad y_{\min} \leq Q^{\dagger} v_k \leq y_{\max} \\ &: \quad y_{\min} \leq Q^{\dagger}\{F^{-1} \begin{bmatrix} t_{k+1} \\ t_{k+2} \\ \vdots \\ t_{k+P} \end{bmatrix}\} \leq y_{\max} \end{aligned}$$

where F is the nonlinear mapping between v and t. Thus for the Wiener models the output constraints are nonlinear in terms of the latent variables. These output constraints can be included in the optimization step along with the input constraints.

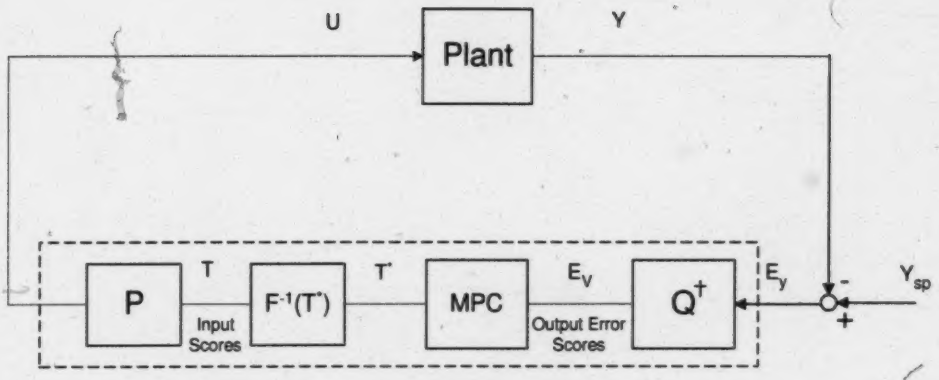


Figure 4.19: *Control strategy for PLS based Hammerstein models*

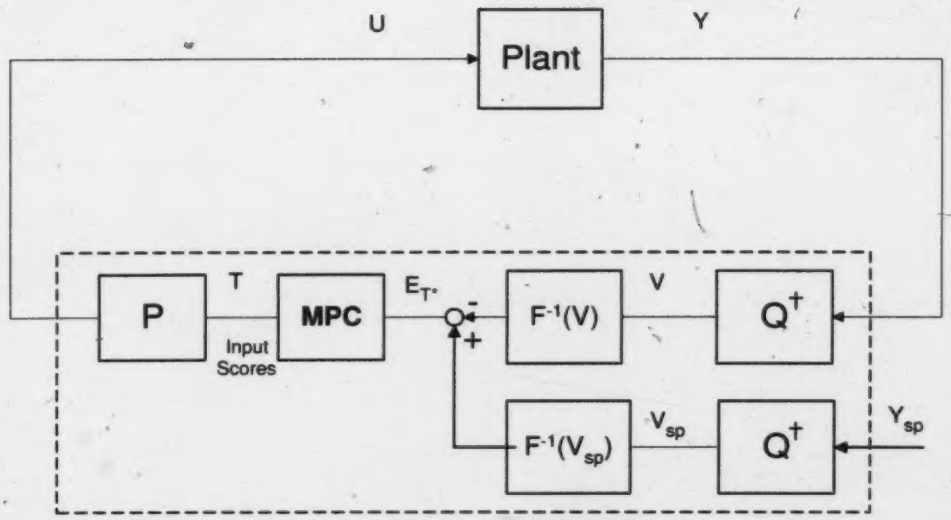


Figure 4.20: *Control Strategy for PLS based Wiener Models*

PLS is generally associated with dimensionality reduction. The main purpose of this work is not dimensionality reduction but to exploit the structural advantages of PLS based models to do constrained control of a class of nonlinear systems. PLS is mainly used here to facilitate transformation of variables to a new basis space where the relationship between the input and output scores is decoupled.

In cases where there is dimensionality reduction the PLS model captures a large portion of the covariance between the inputs and outputs. In such cases the lower dimensional PLS model will still capture the dominant dynamics of the process. The dimensionality reduction will, however, affect the constraint mappings. When there is a dimensionality reduction the P and/or Q matrices will be, in general, non-square. This will lead to a loss in information when the constraints are mapped from the original space to the latent space. But this problem can be resolved by using the full P and Q matrices (without dimensionality reduction) for constraint mapping alone.

4.4.4 Case Study: pH-Level Control of an Acid-Base Neutralization process

The proposed approach was evaluated on a MIMO, pH-level control process. This example is an acid-base (HCl-NaOH) neutralization process performed in a single tank. pH neutralization processes are challenging control problems due to the inherent nonlinearity in the titration curve. The system description, the nonlinear process model and the operating conditions are described in Henson and Seborg (1994). The level and the pH are the controlled variables that are manipulated using the acid and the base flow rates. The buffer flow rate serves as a disturbance. The nominal operating conditions of the process are:

$$\begin{aligned} \text{Outputs} &: \text{pH} = 7.075,\ \text{level} = 14cm \\ \text{Inputs} &: \text{acid} = 16.6\ ml/s,\ \text{base} = 15.6\ ml/s,\ \text{buffer} = 0.6\ ml/s \end{aligned}$$

Control of this process was attempted using a linear PLS based MPC and compared with the nonlinear MPC approach based on Wiener/Hammerstein models as proposed here. A simulation model of the process was built in the Simulink/Matlab environment. The open loop simulations shown in figure 4.21 indicate the nonlinear nature of the pH response. Though the pH response is fairly nonlinear in terms of the gains, it should be noted that the dynamics remain fairly unchanged. It is this feature that makes the pH system suitable for modeling using Hammerstein and Wiener models which have a static nonlinear component. The identification experiments have to be designed to capture the nonlinearity of the system

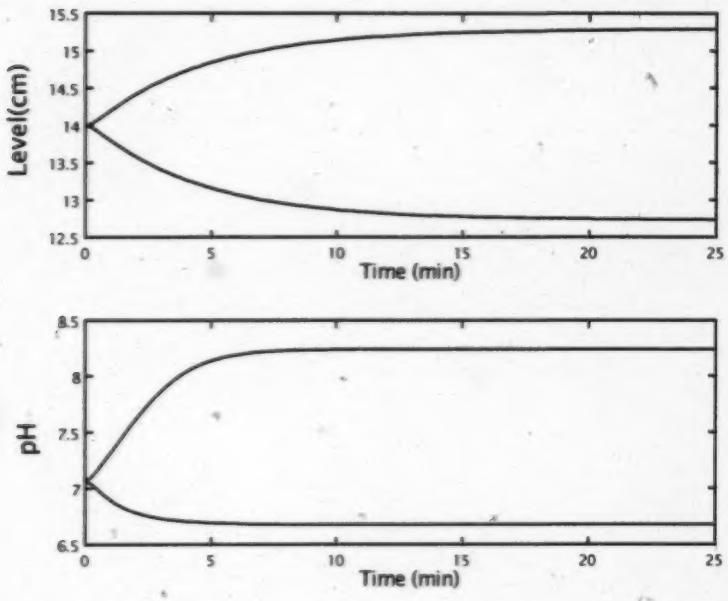


Figure 4.21: *Open-loop simulation: Level and pH responses to step changes in acid flow rate (± 1 ml/s)*

i.e., the magnitude of the input changes is of prime importance and will decide what kind of nonlinearities are excited. To capture the nonlinearity in the gain, the probing signals were designed to be low frequency excitation spread over several levels of magnitude.

PLS based, linear and nonlinear, models were estimated for the system using data collected from the identification runs. The model parameters are given below. Two PLS dimensions ($n = 2$) were used to explain the information in the data set. Thus there was no dimensionality reduction for this example. A fifth order polynomial was required for the Hammerstein model whereas a third order polynomial proved sufficient for the Wiener model. The inner relationship for the second PLS dimension was linear for both methods.

Linear Model:

The variables are normalized (unit variance, zero mean) by the scaling matrices:

$$S_u = diag(1.135, 1.0122), S_y = diag(1.2875, 1.3079).$$

The loading matrices are:

$$P = \begin{bmatrix} 0.7220 & 0.6796 \\ -0.6919 & 0.7335 \end{bmatrix}, Q = \begin{bmatrix} 0.0347 & 1 \\ -0.9994 & 0.0081 \end{bmatrix}.$$

The linear dynamic parts were estimated as:

$$G_1 = \frac{0.0267z^{-1} - 0.0094z^{-2}}{1 - 1.5885z^{-1} + 0.006z^{-2}}, G_2 = \frac{0.0435z^{-1} + 0.0539z^{-2}}{1 - 0.8794z^{-1} - 0.0556z^{-2}}.$$

Hammerstein Model:

Scaling Matrices: The variables are normalized (unit variance, zero mean) by the scaling matrices:

$$S_u = diag(1.135, 1.0122), S_y = diag(1.2875, 1.3079).$$

The loading matrices are:

$$P = \begin{bmatrix} 0.7220 & 0.6796 \\ -0.6919 & 0.7335 \end{bmatrix}, Q = \begin{bmatrix} 0.0347 & 1 \\ -0.9994 & 0.0081 \end{bmatrix}.$$

The static nonlinear part is given by:

$$\begin{aligned} t_1^* &= 0.025t_1^5 + 0.1227t_1^4 - 0.0978t_1^3 - 0.5909t_1^2 + t_1 \\ t_2^* &= t_2. \end{aligned}$$

The linear dynamic parts were estimated as:

$$G_1 = \frac{0.161z^{-1} - 0.180z^{-2}}{1 - 0.8849z^{-1} + 0.0388z^{-2}}, G_2 = \frac{0.0458z^{-1} + 0.0522z^{-2}}{1 - 0.8744z^{-1} - 0.0601z^{-2}}.$$

The cross-validation of the Hammerstein model is shown in figure 4.22. It is compared with linear PLS based model. As can be seen from the figure, the Hammerstein model gives a good fit for the pH whereas the linear model fails to capture the inherent nonlinearity for the pH, but does well with the level part.

Wiener Model

The scalings chosen for the Wiener model were:

$$S_u = diag(1.1989, 1.0202), S_y = diag(1.0664, 1.0252).$$

The loading matrices are given by:

$$P = \begin{bmatrix} 0.6806 & 0.7103 \\ -0.7334 & 0.7039 \end{bmatrix}, Q = \begin{bmatrix} 0.0665 & 0.9981 \\ -0.9978 & -0.0611 \end{bmatrix}.$$

The linear dynamic parts were estimated as:

$$G_1 = \frac{0.1740z^{-1}}{1 - 0.8754z^{-1}}, G_2 = \frac{0.1487z^{-1}}{1 - 0.432z^{-1} - 0.484z^{-2}}.$$

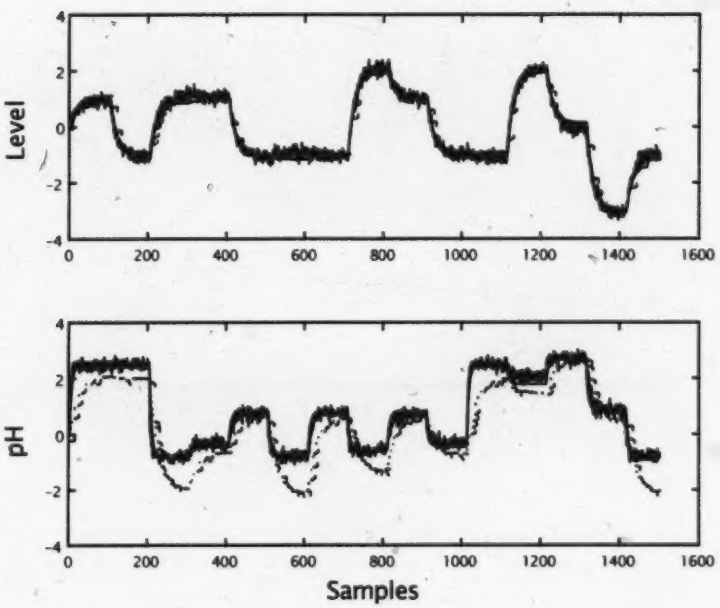


Figure 4.22: *Cross-Validation: Comparison of the Hammerstein model(- -) with a linear model (..) and the actual data (-)*

The static nonlinear parts were given by:

$$v_1 = -0.1173t_1^{*3} - 0.3539t_1^{*2} + t_1^*$$
$$v_2 = t_2^*.$$

The cross-validation plots for the Wiener model are shown in figure 4.23. The Wiener model gives a good fit for the pH part and appears to have captured the nonlinearity present in the system.

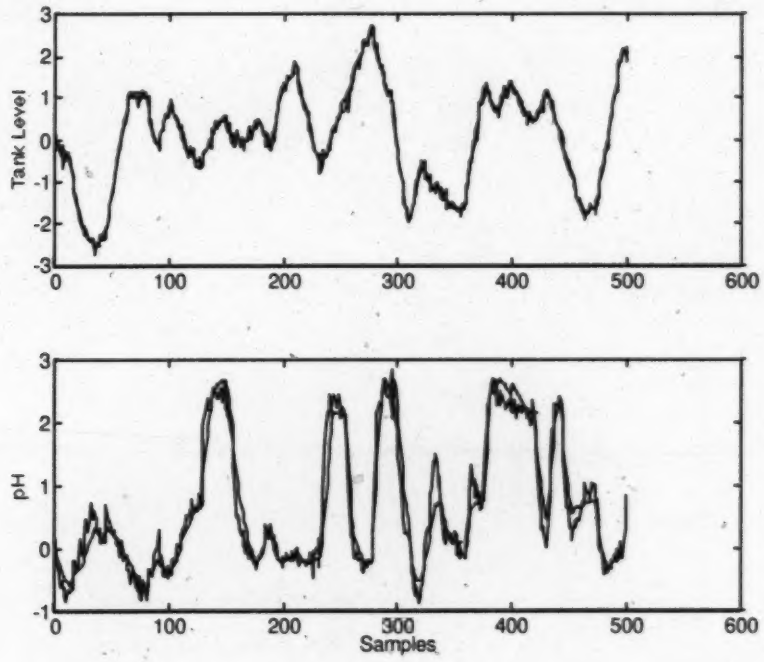


Figure 4.23: *Cross-validation fit for the Wiener model*

Three PLS based controllers were evaluated - (1) Linear MPC, (2) Hammerstein based NLMPC, and (3) Wiener based NLMPC. For control schemes (1) and (3), the Matlab function *qp* was used to solve the quadratic program whereas for (2) the nonlinear programming solver *constr*, available under the Optimization Toolbox in Matlab, was invoked. The sampling time was 15 seconds.

Selection of Controller Tuning Parameters:

$$p = 15, m = 2, \Lambda = 0.5I, \Gamma = I, -1 \leq \Delta u_i(k) \leq 1, -10 \leq u_i(k) \leq 10$$

Figure 4.24 shows the performance of the three controllers in response to a setpoint

change in pH (pH=9). The Wiener based NLMPC gives the best performance amongst the three controllers. The change in pH is achieved in a satisfactory manner with no overshoot in a settling time of 2 minutes. On the other hand the linear MPC showed very poor performance in achieving the pH setpoint change. It appears to struggle with the nonlinearities in the system before reaching the setpoint in 15 minutes. The Hammerstein based NLMPC showed good tracking properties but its disturbance rejection for level was less than satisfactory. As opposed to the linear MPC it gave smooth response in both the channels. Thus both the Hammerstein and Wiener based MPC schemes appeared to annul the nonlinearities in the system through feedback and gave responses similar to a linear system.

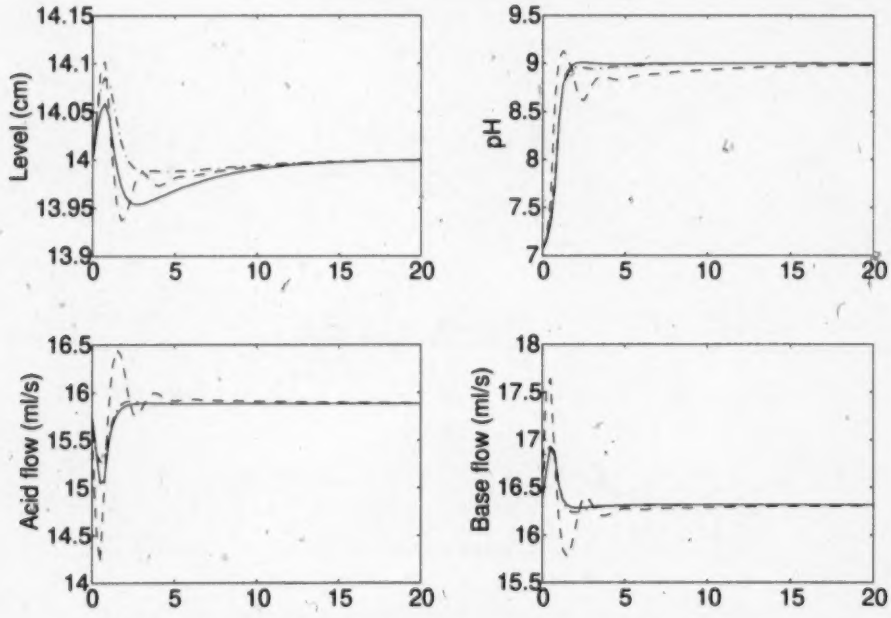


Figure 4.24: *Comparison of (1) Linear MPC (- -), (2) NLMPC-Hammerstein (-) and (3) NLMPC-Wiener (-.) for stepoint changes: pH=9, level=14.*

Figure 4.25 shows the tracking properties of control schemes (2) and (3) for setpoint changes in both level and pH. Clearly the Wiener based NLMPC gives a superior performance. It is able to achieve this setpoint change into a highly nonlinear region in a settling time of 5 minutes. The Hammerstein based NLMPC is able to reach the pH setpoint only if the simulation was continued for long period. This was observed for other setpoint changes as well. One of the reasons for the poor performance of the Hammerstein based NLMPC

could a poor model in the neighborhood of the setpoint. In the Hammerstein model the nonlinearity is at the input and hence the magnitude of input changes plays an important role during the identification process. For example, in figure 4.25 the inputs undergo large changes in magnitude in order to meet the desired setpoint. If the Hammerstein model is unable to capture the nonlinearities associated with these input changes then the performance of the NLMPC will deteriorate. On the other hand for Wiener models the nonlinearity is at the output and therefore the range of magnitude changes at the output affects the goodness of a Wiener model. From figure 4.25 we see that the pH changes are not large in magnitude and the experiment design included changes of this magnitude. Moreover pH is defined as the negative of the natural logarithm of hydrogen ion concentration. This implies that there is a significant nonlinearity at the output. Thus a Wiener structure is naturally suited to capture the nonlinearity of this pH process. The importance of appropriate magnitude changes in the input excitation is illustrated through this result. *A priori* knowledge of the operating conditions of the plant should be taken into account while designing the input sequence for identification.

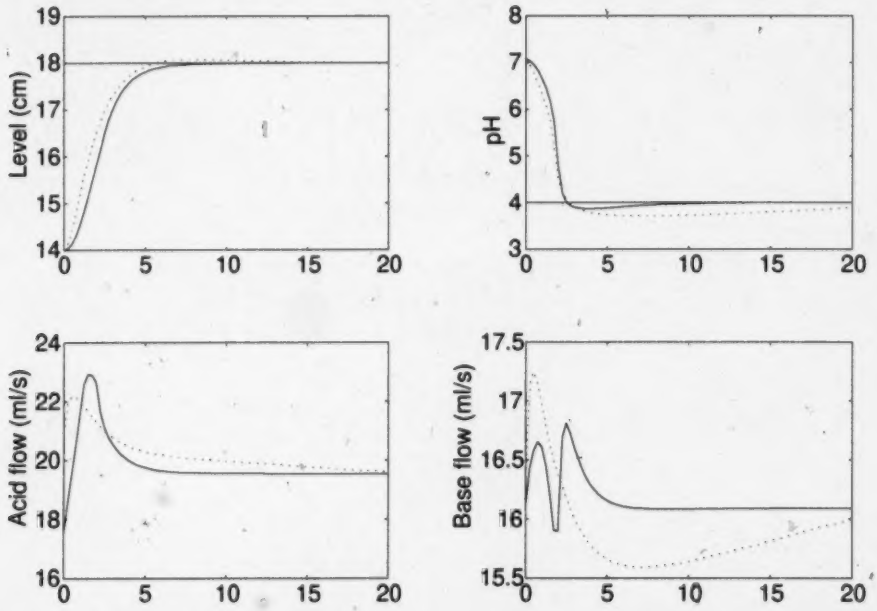


Figure 4.25: *Tracking performance of NLMPC based on (1) Hammerstein models (..) and (2) Wiener models (-) for stepoint changes: pH=4, level=18.*

The disturbance rejection properties of the three controllers are illustrated in figure 4.26.

Once again Hammerstein and Wiener based MPC schemes outperform the linear MPC. The main nonlinearity of the process lies in the pH part and the linear MPC does not do a good job of dealing with it. The integrator in the linear MPC, however, does take the pH back to its steady state value.

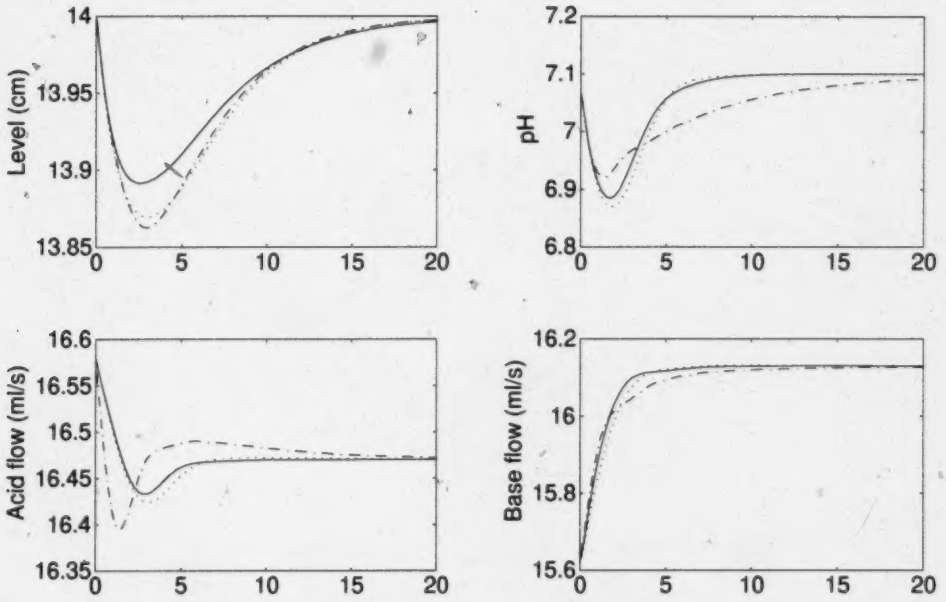


Figure 4.26: *Response of (1) Linear MPC (-.), (2) NLMPC-Hammerstein (..) and (3) NLMPC-Wiener (-) to a change in buffer flow (0.6 to 0.2).*

The Wiener-NLMPC was able to meet several setpoint changes into nonlinear regions which the other controllers could not handle. Figure 4.27 shows the performance of the Wiener-NLMPC and figure 4.28 shows the constraints being satisfied in the original domain and the equivalent constraints in the latent space. To better illustrate the constraint satisfaction and the geometry, the upper and lower limits on the inputs and the slew rates were reduced for this run:

$$-0.5 < \Delta u_i < 0.5, -5 < u_i < 5; i = 1, 2.$$

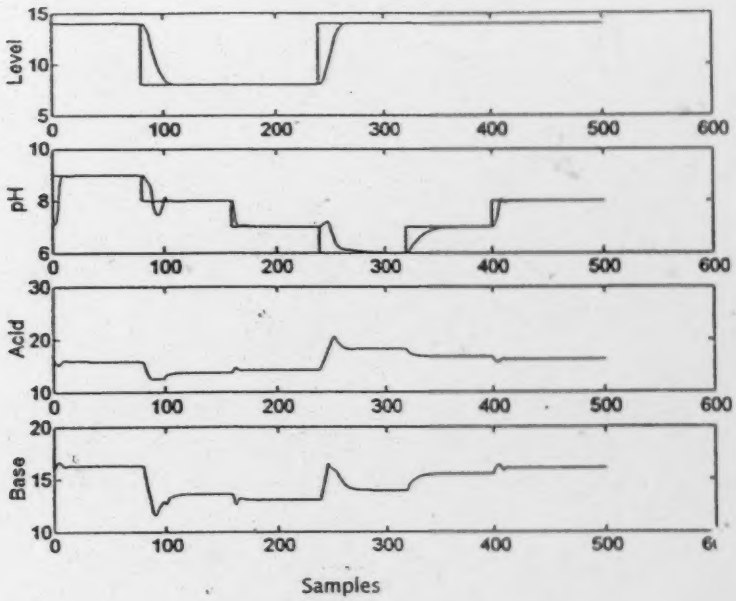


Figure 4.27: *Performance of Wiener based NLMPC for a series of setpoint changes.*

Discussion

A method of handling constraints for nonlinear model predictive control using Hammerstein and Wiener structures in the PLS framework is presented in this paper. The feasibility of the approach is demonstrated via a simulation case study of a MIMO system involving pH and level control. The main contributions of this work are (1) a technique to effectively use Hammerstein and Wiener structures for modeling and control of MIMO systems via the PLS framework and (2) constrained NLMPC of nonlinear processes using these models. The main advantage of using the PLS framework is that the dynamics are decoupled in the latent space. The constraints, however, get coupled in the latent space. Thus a multivariable approach to controller synthesis has to be adopted for constrained control. For Hammerstein models, the constraints are mapped onto a nonlinear region in the latent space; for Wiener models the constraints, though coupled in terms of the latent variables, maintain their linearity.

For Wiener models it is shown that only a quadratic program needs to be solved, which is

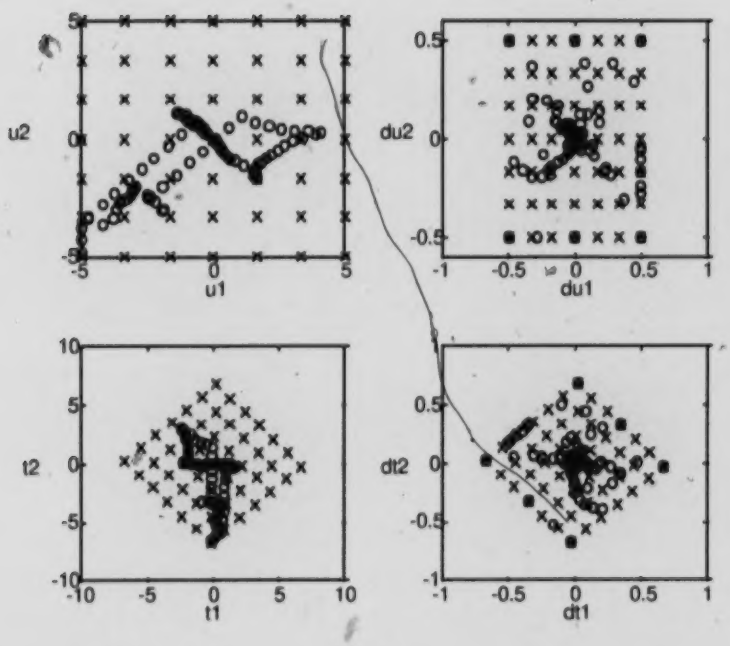


Figure 4.28: *Rate and Amplitude Constraints in the original space (top graph) and the latent space (bottom graph). 'x' - denotes the constraint region and 'o' denotes actual data.*

a significant computational advantage over other NLMPC methods and makes them suitable for controller design purposes. Since block oriented models serve as useful representations for chemical processes such as distillation and acid-base neutralization (Norquay *et al.*, 1996), it is worthwhile to consider modeling and control of MIMO processes using these structures.

The major advantage of the use of PLS based Hammerstein and Wiener models is their ability to decouple the dynamics and the nonlinearities. Even though the presence of constraints negates this advantage, it is important to note that the numerical inversion of the static nonlinearities is restricted to a single variable root finding. In the absence of the PLS structure, the root finding operation would be a multivariable one and is likely to present numerical difficulties.

4.5 Conclusions

In this chapter we considered PLS based dynamic models for:

1. **Interactor estimation and multivariable time delay estimation.** Several attractive features of the PLS model form were emphasized. The PLS model was shown to be similar to a cannonical form identified from input-output data. The interactor matrix can be easily estimated and multivariable time delay compensation requires simply requires n univariate Smith predictors in the latent space.

2. **Constrained model predictive control in the latent space.** Constrained MPC was formulated in terms of the latent inputs and outputs. The presence of constraints leads to coupling between the latent inputs. For linear processes, this approach was demonstrated on a simulation of the Wood-Berry column and a pilot scale stirred tank heater. A class of nonlinear systems described by Hammerstein and Wiener models was also considered via the PLS based MPC. A pH-level control problem demonstrated the applicability of the nonlinear PLS based MPC.

Chapter 5

Techniques for assessment of model predictive controllers

5.1 Introduction

The area of performance assessment is concerned with the performance analysis of existing controllers. Performance assessment aims at evaluating controller performance from routine data. Several algorithms are now available for estimating a performance index from closed loop data. Conventionally the estimation procedure involves comparison of the existing controller with a theoretical benchmark such as the minimum variance controller (MVC). Harris (1989) laid the theoretical foundations for performance assessment of single loop controllers from routine operating data. Time series analysis of the output error was used to determine the minimum variance control for the process. A comparison of the output variance term with the minimum achievable variance reveals how well the current controller is doing currently. Subsequently Huang *et al.* (1997) and Harris *et al.* (1997) extended this idea to the multivariate case based on the generalization of the time delay term. The multivariate interpretation of the delay term is known as the *interactor matrix* and this entity plays a crucial role in determining the control invariant or the minimum variance control for a MIMO process. However, estimation of the interactor matrix requires some closed loop excitation and cannot be based on routine operating data alone. Huang (1997) showed that the interactor matrix obtained under closed loop conditions is the same as the open loop interactor. Kozub and Garcia (1993) proposed user defined benchmarks based on settling times, rise times, *etc.* This presents a more practical method of assessing controller performance. The settling time or rise time for a process can often be chosen based on operator experience. A correlation analysis of the operating data is used to determine whether

[1]A version of this chapter was presented as "Performance Assessment of Model Predictive Controllers", 1997, Rohit S. Patwardhan, Biao Huang and Sirish L. Shah, CSChE Annual Meeting, Edmonton, Canada.

the desired closed loop characteristics were achieved. Tyler and Morari (1995) proposed performance evaluation based on likelihood methods and hypothesis testing. Performance assessment of non-minimum phase and open loop unstable systems was also addressed by Tyler and Morari (1995).

Huang (1997) also proposed the LQG control as the benchmark instead of MVC. The main advantage of this technique is that the input variance is also taken into account. In many processes the input variance is of major concern as this is often a utility such as steam, power, *etc.* with significant cost. A model of the process and the disturbances is required to do the LQG benchmarking. Kammer *et al.* (1996) used non-parametric modeling in the frequency domain to ascertain the optimality of a LQG controller, based on the comparison of the optimal and the achieved cost functions.

Model predictive controllers belong to a class of model based controllers which compute future control actions by minimizing a performance objective function over a finite prediction horizon. This family of controllers is truly multivariate in nature and has the ability to deal with constraints on the inputs, slew rates, *etc.* It is for the above reasons that MPC has been widely accepted by the process industry. Variations of MPC such as dynamic matrix control (DMC) and quadratic dynamic matrix control (QDMC) have become the norm in industry for processes where interactions are of foremost importance and constraints have to be taken into account.

Performance assessment of MPC presents a challenging research problem. A constrained MPC type controller is essentially a nonlinear controller, especially when operating at the constraints. Conventional MVC benchmarking is infeasible and alternative techniques have to be developed. Two such methods for MPC performance evaluation are proposed here - (1) the use of LQG as a benchmark based on the knowledge of the open loop process and noise models, and (2) use of the design performance as a benchmark.

The rest of this chapter is organized as follows: Section 5.2 gives a tutorial introduction to the conventional performance assessment techniques based on minimum variance benchmarking and closed loop settling time specifications. In Section 5.3 we propose the use of the design case as a benchmark for model predictive controllers. A performance measure based on comparison of the design objective function with the achieved objective function is defined and some properties of this measure are established. The design case benchmark and the LQG benchmark are applied to evaluate the MPC performance for two simulated processes in Section 5.4. The effect of constraints is highlighted. An industrial QDMC application on a recycle surge drum level control is used to demonstrate the practical na-

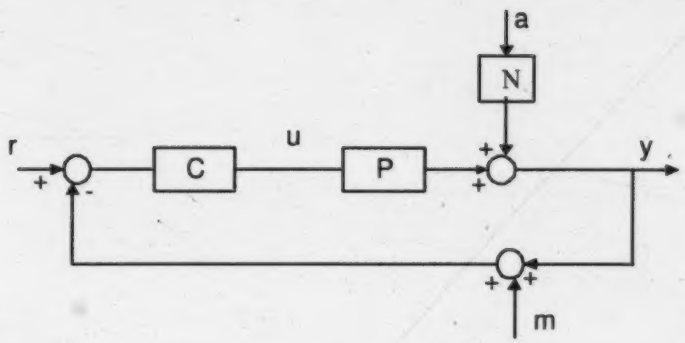


Figure 5.1: *A conventional control system with unmeasured disturbances and measurement noise*

ture of the performance measures, in Section 5.5. The contributions of this chapter are summarized in Section 5.6.

5.2 Conventional Assessment Techniques

5.2.1 Minimum Variance Benchmarking

Univariate Case

Minimum variance benchmarking is based on the concept, that under minimum variance control the closed loop response will exactly match the setpoint change or cancel the disturbance after d samples, where d is the process time delay. Consequently it follows that the variance of a closed loop system within the first d samples is controller invariant and corresponds to the minimum achievable variance. For a rigorous proof see Harris (1989) or Huang (1997). Consider the following closed loop system.

The closed loop relationship between the unmeasured disturbances and the process output is given by

$$y(k) = \frac{N}{1+PC}a(k) = \frac{N}{1+z^{-d}\tilde{P}C}$$

where d is the process time delay. The above relationship can also be expressed as an infinite-order moving average (MA) process, for closed loop stable systems,

$$y(k) = (f_0 + f_1 z^{-1} + \ldots + f_{d-1} z^{-d+1} + f_d z^{-d} + \ldots)a(k).$$

The first d terms define the minimum achievable variance,

$$\sigma^2_{mv} = (f_0^2 + f_1^2 + \ldots + f_{d-1}^2)\sigma_a^2.$$

The controller performance index is defined as the ratio of the minimum achievable variance to the achieved output variance,

$$0 \leq \eta(d) = \frac{\sigma^2_{mv}}{\sigma^2_y} \leq 1.$$

If the performance index is equal to one then it implies that the process is under minimum variance control. On the other hand if the index assumes the value of zero it implies that the process is unstable under closed loop. Desborough and Harris (1992, 1993) give statistical bounds on this performance measure and also extend it to assess feedforward control. Theoretically this measure can also be extended to the case where there are setpoint changes. In such cases, the output error can be expressed as a infinite order moving average model. Huang (1997) gives the basis for extending this method to the case of deterministic disturbances/setpoints.

Example 5.1. *This is a simulation example from Desborough and Harris (1992), with time delay, d=2:*

$$y(k) = u(k-2) + \frac{1-0.2z^{-1}}{1-z^{-1}}a(k).$$

A simple integral controller was used for control, $\Delta u(k) = -Ky(k)$. For K=0.5, the closed loop response is given by:

$$y(k) = \frac{1-0.2z^{-1}}{1-z^{-1}+0.5z^{-1}}a(k)$$
$$= (1+0.8z^{-1}+0.3z^{-2}-0.1z^{-3}-0.25z^{-4}+...)a(k).$$

For K=0.2, we have

$$y(k) = \frac{1-0.2z^{-1}}{1-z^{-1}+0.2z^{-1}}a(k)$$
$$= (1+0.8z^{-1}+0.6z^{-2}+0.44z^{-3}+0.32z^{-4}+...)a(k).$$

Notice that the first two terms in each case are identical. In fact they will be the same irrespective of the controller used. The coefficients of the MA model, for the two proportional gains are plotted in the figure 5.2. The performance indices for the two controllers can be estimated by comparing the minimum variance with the achieved variance. This concept is illustrated by the shaded areas in the figure 5.3 .

	K=0.5	*K=0.2*
$\eta(d)$	*0.8803*	*0.6808*

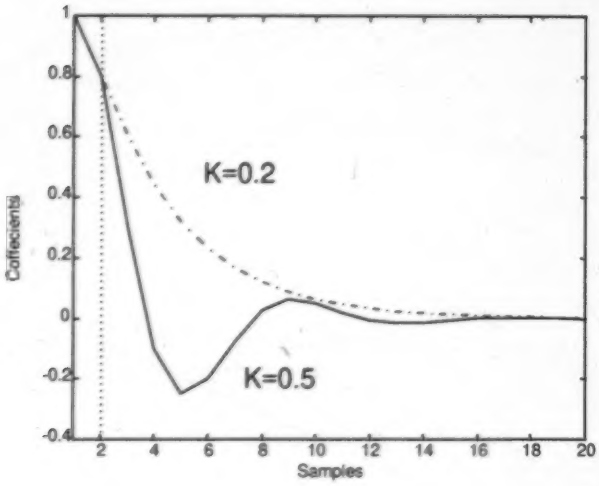


Figure 5.2: *The closed loop impulse response coefficients for the two controller gains.*

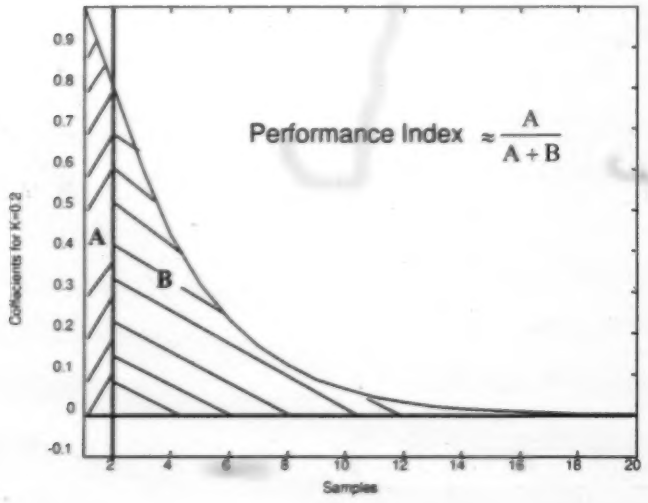


Figure 5.3: *A graphical illustration of the performance index calculation.*

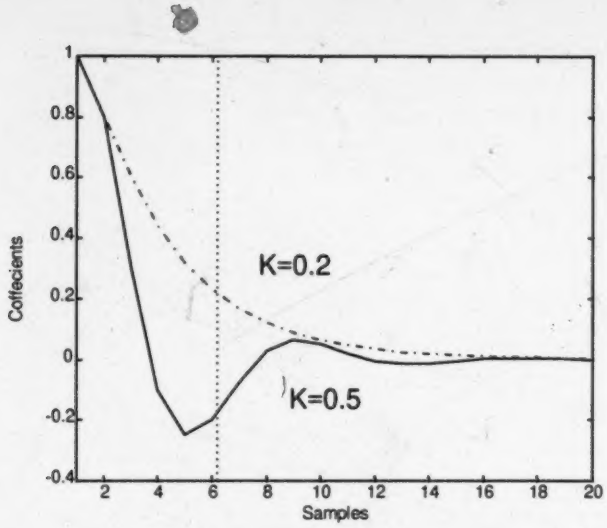


Figure 5.4: *Illustration of the desired settling time based benchmark.*

Kozub and Garcia (1993) proposed the use of desired closed loop settling time as a benchmark of performance. We demonstrate this idea using the above example. Let us fix the desired closed loop settling time as 6 samples i.e. the output variance should ideally reduce to zero after size samples. The performance index is then defined as

$$\xi(t_c) = \frac{\sigma^2_{des}}{\sigma^2_y}$$

where $\sigma^2_{des} = (f_0^2 + f_1^2 + \ldots + f^2_{t_c-1})\sigma^2_a$. *The corresponding performance indices for the two controller gains are given below.*

	K=0.5	*K=0.2*
$\xi(t_c)$	*0.9927*	*0.9754*

Thus in this case both the controllers are achieving the design performance according to the settling time based measure. It is clear from the figure 5.4 that this is a controller dependent measure.

Multivariable Minimum Variance Benchmarking

Harris *et al.* (1997) and Huang *et al.* (1997) extended the idea of minimum variance benchmarking to the multivariable case based on the knowledge of the interactor matrix. First we state the result from Huang (1997) which proves the existence of the multivariable minimum variance controller.

Theorem 5.1. *For a multivariable process*

$$y = Pu + Na$$

with the linear quadratic objective function (singular LQ objective function) defined by

$$J = E(\tilde{y}^T \tilde{y})$$

where $\tilde{y} = z^{-d}Dy$, *an explicit optimal solution is given by*

$$u = -\tilde{P}^{-1}RM_F^{-1}Dy = -\tilde{P}^{-1}RF^{-1}(z^{-d}D)y$$

where $\tilde{P} = DP, M_F = z^d F, F$ *and* R *satisfy the Diophantine identity:*

$$z^{-d}DN = \underbrace{F_0 + ... + F_{d-1}z^{-d+1}} + z^{-d}R$$

where R *is a rational proper transfer function matrix and* D *is the unitary interactor.*

The key steps involved in the multivariable minimum variance benchmarking algorithm due to Huang (1997) are:

1. Filter the process outputs (errors) with the interactor. The interactor can be obtained through a model of the process or from closed loop identification or through some prior knowledge about the delays in each channel,

$$\tilde{y}(k) = (D_0 + D_1 z^{-1} + ... + D_d z^{-d})y(k).$$

2. The filtered output now has a simple interactor, and the minimum variance benchmarking can be applied in a manner quite similar to the SISO case. First the outputs (errors) are pre-whitened using a multivariable auto-regressive (AR) model.

$$(A_0 + A_1 z^{-1} + ... + A_{nw} z^{-nw})\tilde{y}(k) = e(k) \Leftrightarrow$$
$$(B_0 + B_1 z^{-1} + ... + B_d z^{-d} + ...)e(k) = \tilde{y}(k)$$

   where e is a white noise process.

3. Multivariable cross-correlation analysis between the estimated white noise sequence and the interactor filtered outputs gives the covariance matrix under minimum variance control,

$$\sigma_{mv}^2 = (B_0 \sigma_e^2 B_0^T + B_1 \sigma_e^2 B_1^T + ... + B_{d-1} \sigma_e^2 B_{d-1}^T).$$

4. The multivariable performance index is then obtained as

$$\eta = \frac{tr(\sigma_{mv}^2)}{tr(\sigma_y^2)}.$$

where tr denotes the trace of the matrix[2].

In mere words, we are taking the ratio, of the sums of individual variances under MVC, to the achieved variance. For a discussion of the theoretical issues involved in interactor estimation and its application to minimum variance benchmarking see Huang (1997). See Huang (1997) for an application of multivariable minimum variance benchmarking to a 2x2 simulation example.

LQG benchmarking

The linear quadratic Gaussian control law offers the best achievable control for a linear plant. The minimum variance control law can be derived as a special case of the LQG controller. For the regulatory case the infinite horizon LQG objective function can be written as:

$$\begin{aligned} J_{lq} &= \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{k=1}^{N} \hat{y}(k)^T \Gamma \hat{y}(k) + \Delta u(k-1)^T \Lambda \Delta u(k-1)\right\} \\ \hat{x}(k+1) &= A\hat{x}(k) + B_1 \Delta u(k) \\ y(k) &= C\hat{x}(k) \end{aligned}$$

where E denotes the expected value. Here integral action has been included into the LQG control law and the objective function is in the input-output form. The optimal linear feedback control law can be derived through minimization of the above objective function,

$$\Delta u(k) = -K\hat{x}(k) \tag{5.1}$$

where

$$K = (R + B_1^T S B_1)^{-1} B_1^T S A.$$

S is the solution of the steady state algebraic Riccati equation:

$$S = A^T \left[S - PB_1(\Lambda + B_1^T S B_1)^{-1} B_1^T S\right] A + \Gamma.$$

The optimal state estimator or the Kalman filter completes the LQG feedback law,

$$\hat{x}(k+1) = A\hat{x}(k) + B_1 \Delta u(k) + L(y(k) - \hat{y}(k)).$$

[2]The trace of matrix is defined by $tr(A) = \sum_{i=1}^{n} a_{ii}$ i.e. the sum of its diagonal elements.

If the true process is described by

$$\begin{aligned} x(k+1) &= Ax(k) + B\Delta u(k) + Gw(k) \\ y(k) &= Cx(k) + v(k) \end{aligned}$$

where $w(k), v(k)$ are state and measurements noises with known covariance matrices $E(ww^T) = \sigma_w^2, E(vv^T) = \sigma_v^2, E(wv^T) = \sigma_{wv}^2$. The resulting closed loop system of equations can be written as

$$\begin{aligned} \begin{bmatrix} x(k+1) \\ \hat{x}(k+1) \end{bmatrix} &= \begin{bmatrix} A & -BK \\ LC & A - BK - LC \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \end{bmatrix} + \begin{bmatrix} G & 0 \\ 0 & L \end{bmatrix} \begin{bmatrix} w(k) \\ v(k) \end{bmatrix} \\ \begin{bmatrix} \Gamma^{1/2} y(k) \\ \Lambda^{1/2} \Delta u(k) \end{bmatrix} &= diag(\Gamma^{1/2}\ \Lambda^{1/2}) \left\{ \begin{bmatrix} C & 0 \\ 0 & -K \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \end{bmatrix} + \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w(k) \\ v(k) \end{bmatrix} \right\} \end{aligned}$$

In terms of new state and output variables we have,

$$\begin{aligned} \eta(k+1) &= A_c\eta(k) + B_c\varsigma(k) \\ \psi(k) &= C_c\eta(k) + D_c\varsigma(k). \end{aligned} \tag{5.2}$$

If $w(k)$ and $v(k)$ are uncorrelated white noise sequences then it can be shown that $E(J_{LQG})$ is the H_2 norm[3] of the closed loop system in equation 5.2,

$$\begin{aligned} G &\triangleq \left[\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right] \\ ||G||_2^2 &= tr(D_cD_c^T + C_cPC_c^T) \end{aligned}$$

where P is the solution to the Lyapunov equation, $P = A_cPA_c^T + B_cB_c^T$ (Chen and Francis, 1995). For the case where $w(k)$ and $v(k)$ are correlated, the optimal cost function can be found by solving relevant Lyapunov equations (see Huang, 1997).

The LQG performance curve in figure 5.5 is a graph of the design output variance against the design input variance for different control weightings (Λ). The LQG performance index is a measure of the distance of any controller from the achievable bound. A comparison of the achieved objective function with the LQ objective yields one such measure:

$$\begin{aligned} \rho &= \frac{J_{lq}}{J} \\ J &= E\{\frac{1}{N}\sum_{k=1}^{N} y(k)^T\Gamma y(k) + \Delta u(k-1)^T\Lambda\Delta u(k-1)\}. \end{aligned} \tag{5.3}$$

The LQG benchmark provides useful information about the re-tuning of a model based

[3] 2-norm: $||P||_2 = (\frac{1}{2\pi}\int_0^{2\pi} |P(e^{jw})|^2 dw)^{1/2} = P(0)^2 + P(1)^2 + ...$

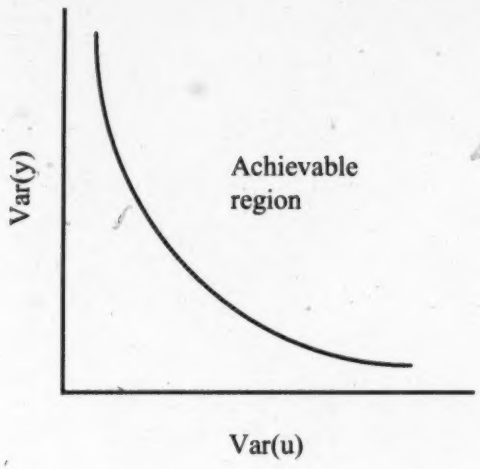


Figure 5.5: *The optimal performance curve obtained through LQG benchmarking*

predictive controller given its similarity in structure and form to a MPC controller. The weightings for the LQG criterion can be chosen to reflect the MPC objective. The LQG approach is more practical compared to MVC since it gives importance to the input variance. The performance curve is sensitive to the ratio of the input to measurement noise variances and the inexact knowledge of this quantity can result in inaccurate benchmarking. Uncertainty in the process and noise models also results in an uncertain performance curve. A low performance index does not always imply poor performance since it is based on comparison of two different control laws, for example, MPC with LQG. The model predictive controller may be delivering the nominal performance it was designed for and yet be far away from the achievable performance curve. The LQG controller is an unconstrained controller and provides a useful lower bound on the achievable performance of a model based predictive controller.

5.3 Performance Assessment of Model Predictive Controllers

In this Section we will restrict ourselves to cases where a MPC type controller is in place and the plant dynamics are linear. Conventionally a linear feedback controller is in place and time series analysis of the closed loop data (output) can be used to evaluate the controller performance. However, MPC gives a linear time-varying feedback law for the constrained case. Evaluation of the MVC benchmark from the nonlinear closed loop data without resorting to identification techniques is a challenging research problem (Harris, 1989). It should

be noted here that for a linear plant, LQG offers the best achievable nominal performance amongst the class of all stabilizing, linear or nonlinear, controllers.

Using the design case as a benchmark

An alternative approach is to evaluate the controller performance using a criterion commensurate with the actual design objective(s) of the controller and then compare the achieved performance. This idea is analogous to the method of Kammer *et al.* (1996) which was based on frequency domain comparison of the achieved and design objective functions for LQG. For a MPC controller with a quadratic objective function, the design requirements are quantified by:

$$\hat{J}_k = (r_k - \hat{y}_k)^T \Gamma (r_k - \hat{y}_k) + \Delta u_k^T \Lambda \Delta u_k.$$

The model predictive controller calculates the optimal control moves by minimizing this objective function over the feasible control moves. If we denote the optimal control moves by Δu_k^*, the optimal value of the design objective function is given by

$$\hat{J}_k^* = \hat{J}_k(\Delta u_k^*).$$

The actual output may differ significantly from the predicted output due to inadequacy of the model structure, nonlinearities, modeling uncertainty, *etc.* Thus the achieved objective function is given by:

$$J_k = (r_k - y_k)^T \Gamma (r_k - y_k) + \Delta u_k^T \Lambda \Delta u_k.$$

where $y_k, \Delta u_k$ denote the measured values of the outputs and inputs at corresponding sampling instants appropriately vectorized. The inputs will differ from the design value in part due to the receding horizon nature of the MPC control law. The value of the achieved objective function cannot be known *a priori*, but only p sampling instants later. A simple measure of performance can then be obtained by taking a ratio of the design and the achieved objective functions,

$$\eta(k) = \frac{\hat{J}_k^*}{J_k}. \tag{5.4}$$

This performance index will be equal to one when the achieved performance meets the design requirements. The advantage of using the design criterion for the purpose of performance assessment is that it is a measure of the deviation of the controller performance from the expected or design performance. Thus a low performance index truly indicates changes in

the process or the presence of disturbances, resulting in sub-optimal control. The estimation of such a index does not involve any time series analysis or identification. The design objective is calculated by the controller at every instant and only the measured input and output data is needed to find the achieved performance. The above performance measure represents an instantaneous measure of performance and can be driven by the unmeasured disturbances. In order to get a better overall picture the following measure is recommended:

$$\alpha(k) = \frac{\sum_{i=1}^{k} \hat{J}_i^*}{\sum_{i=1}^{k} J_i}. \tag{5.5}$$

$\alpha(k)$ is the ratio of the average design performance to the average achieved performance up to the current sampling instant. Thus $\alpha(k) = 1$ implies that the design performance is being achieved on an average. $\alpha(k) < 1$ means that the achieved performance is worse than the design.

$\hat{J}_k^*$ can be obtained from the optimization step in MPC and J_k can be found from the operating data and the controller tuning parameters. Consider the quadratic form which is present in the objective function of MPC. If we let

$$\begin{aligned} J_k(y) &= (r_k - y_k)^T \Gamma (r_k - y_k) \\ J_k(u) &= \Delta u_k^T \Lambda \Delta u_k \end{aligned} \tag{5.6}$$

then we have,

$$\begin{aligned} J_k &= J_k(y_k) + J_k(u_k) \\ ||\mathrm{J}_k|| &= ||\mathrm{J}_k(y_k)|| + ||\mathrm{J}_k(u_k)|| \end{aligned}$$

where $||.||$ is the 1-norm as defined before and $\mathrm{J}_k(y_k)$,$\mathrm{J}_k(u_k)$ are the respective vectors of $[J_i(y_i)]^T, [J_i(u_i)]^T, i = 1, ..., k$. By analogy we have a similar expression for the design objective term,

$$\begin{aligned} \hat{J}_k^* &= \hat{J}_k^*(\hat{y}_k) + \hat{J}_k^*(u_k^*) \text{ and} \\ ||\mathrm{J}_k^*|| &= ||\mathrm{J}_k^*(\hat{y}_k)|| + ||\mathrm{J}_k^*(u_k^*)||. \end{aligned}$$

Thus we have expressed the respective objective functions as a sum of the contributions from the outputs and the inputs. Comparing each of these terms in the above sum with their achieved counterparts can give an idea of which quantity, in a relative sense contributes the

most to a loss in performance. We also define performance indices based on the individual terms in the objective function.

$$\alpha_y(k) = \frac{||\mathrm{J}_k^*(\hat{y}_k)||}{||\mathrm{J}_k(y_k)||}, \alpha_u(k) = \frac{||\mathrm{J}_k^*(u_k^*)||}{||\mathrm{J}_k(u_k)||}$$

The objective function approach has the following advantages: (i) since it is free of any restrictive assumptions, it can deal with the multivariate and constrained nature of MPC, (ii) it reflects the design criterion and answers the question - whether the controller doing what it was designed for, (iii) it does not involve any estimation and therefore the measure is exact.

5.3.1 Properties

In the ensuing discussion we prove several useful properties of the proposed benchmark based on comparison of the design and achieved objective function. First a analytic way of obtaining the expected value of the design cost function is described. As with the LQG benchmark this involves solution of the relevant Lyapunov equations describing the closed loop relationships. The analytic expression is derived for the case of stochastic and deterministic inputs. The state space description of MPC is utilized in obtaining the desired quantities. The effect of uncertainty on the design cost function is also captured in the state space framework.

An important implementational aspect of MPC is its receding horizon nature. It is shown here that the receding horizon implementation leads to a loss in the design performance as per the cost function (Property 5). However, the receding horizon nature is known to combat disturbance uncertainty and therefore may lead to lower achieved cost than a conventional multi-step implementation.

We attempt to show the bounded nature of the proposed index under the assumption of closed stability (Property 6). The convergence analysis provides insight into the controller relevant terms that may affect the achieved performance.

1. **Stochastic inputs.** Here we derive the expected value of the design cost function for MPC for the case of stochastic inputs described by

$$x(k+1) = Ax(k) + Bu(k) + Gw(k)$$
$$y(k) = Cx(k) + v(k).$$

where $w(k), v(k)$ are uncorrelated white noise processes.

The predictions can be expressed in terms of the future inputs and current state variables as (see Chapter 2)

$$\hat{y}_k = F\hat{x}(k) + S\Delta u_k + d_k$$
$$\text{where } F = \begin{bmatrix} CA & CA^2 & \dots & CA^p \end{bmatrix}^T \tag{5.7}$$

where S is once again the dynamic matrix made up of the step response coefficients. The disturbance estimate is given by

$$\begin{aligned} d_k &= [I...I]^T(y(k) - \hat{y}(k)) \\ &= [I...I]^T C[x(k) - \hat{x}(k)] \\ &= C_1(x(k) - \hat{x}(k)) \end{aligned}$$

where $\hat{x}(k)$ is the state estimate obtained through a Kalman predictor/filter. The state space form of the MPC feedback law is given by (see Chapter 2),

$$\begin{aligned} \Delta u_k &= K(r_k - F\hat{x}(k) - d_k) \\ &= K(r_k - I_p y(k) - (F - [I \dots I]^T C)\hat{x}(k)) \\ &= K_1 e(k) - K_2 \hat{x}(k) \end{aligned}$$
$$\text{where } K_1 = K[I \dots I]^T, K_2 = (F - [I \dots I]^T C). \tag{5.8}$$

The objective function now becomes (see equation 2.28),

$$J = \{r_k - F\hat{x}(k) - S\Delta u_k - d_k\}^T \Gamma \{r_k - F\hat{x}(k) - S\Delta u_k - d_k\} + \Delta u_k^T \Lambda \Delta u_k.$$

For the nominal case the feedback term will correspond to the noise term. Take the regulatory case where $r_k = 0$. Therefore we have

$$\begin{aligned} r_k - F\hat{x}(k) - S\Delta u_k - d_k &= -F\hat{x}(k) - S(K_1 y(k) - K_2 \hat{x}(k)) - C_1(x(k) - \hat{x}(k)) \\ &= (-F + SK_2 + C_1)\hat{x}(k) - (SK_1 C + C_1)x(k) - SK_1 v(k) \\ \phi(k) &\triangleq (-F + SK_2 + C_1)\hat{x}(k) - (SK_1 C + C_1)x(k) - SK_1 v(k) \\ \Delta u_k &= -K_1 y(k) - K_2 \hat{x}(k) \\ &= -K_1 x(k) - K_2 \hat{x}(k) - K_1 v(k) \\ \theta(k) &\triangleq -K_1 x(k) - K_2 \hat{x}(k) - K_1 v(k) \\ \hat{J}(k) &= \phi(k)^T \Gamma \phi(k) + \theta(k)^T \Lambda \theta(k). \end{aligned}$$

The expected value of the objective function can be easily expressed in terms of the variances of the pseudo-outputs defined above,

$$E(\hat{J}(k)) = E[\phi(k)^T \Gamma \phi(k) + \theta(k)^T \Lambda \theta(k)]$$
$$= tr\{\Gamma E[\phi(k)\phi(k)^T]\} + tr\{\Lambda E[\theta(k)\theta(k)^T]\}.$$

Computing $E[\phi(k)\phi(k)^T]$ and $E[\theta(k)\theta(k)^T]$:

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + P[y(k) - \hat{y}(k)]$$
$$\Delta u(k) = -[I\ 0...0](K_1 y(k) + K_2 \hat{x}(k)) = -K_1' y(k) - K_2' \hat{x}(k).$$

For the integrator we define an extra state of the following form:

$$z(k+1) = z(k) + \Delta u(k) = z(k) - K_1' y(k) - K_2' \hat{x}(k)$$
$$u(k) = z(k).$$

Writing the space description together we have

$$\begin{bmatrix} x(k+1) \\ \hat{x}(k+1) \\ z(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 & B \\ PC & A-PC & B \\ -K_1'C & -K_2' & I \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \\ z(k) \end{bmatrix} + \begin{bmatrix} B_1 & 0 \\ 0 & P \\ 0 & -K_1' \end{bmatrix} \begin{bmatrix} w(k) \\ v(k) \end{bmatrix}$$

$$\begin{bmatrix} \phi(k) \\ \theta(k) \end{bmatrix} = \begin{bmatrix} -(SK_1C + C_1) & (-F + SK_2 + C_1) & 0 \\ -K_1 & -K_2 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \\ z(k) \end{bmatrix} + \begin{bmatrix} 0 & -SK_1 \\ 0 & -K_1 \end{bmatrix} \begin{bmatrix} w(k) \\ v(k) \end{bmatrix}$$

or

$$\xi(k+1) = A_F \xi(k) + B_F \psi(k)$$
$$y_J(k) = C_F \xi(k) + D_F \psi(k).$$

The state co-variance matrices can be obtained by solving the equivalent covariance matrices,

$$\xi(k+1)\xi(k+1)^T = (A_F\xi(k) + B_F\psi(k))(A_F\xi(k) + B_F\psi(k))^T \tag{5.9}$$
$$= A_F\xi(k)\xi(k)^T A_F^T + 2A_F\xi(k)\psi(k)^T B_F^T + B_F\psi(k)\psi(k)^T B_F^T$$
$$E[\xi(k+1)\xi(k+1)^T] = A_F E[\xi(k)\xi(k)^T] A_F^T + B_F E[\psi(k)\psi(k)^T] B_F^T$$
$$\sigma_\xi^2 = A_F \sigma_\xi^2 A_F^T + B_F \sigma_\psi^2 B_F^T$$
$$\sigma_{y_J}^2 = C_F \sigma_\xi^2 C_F^T + D_F \sigma_\psi^2 D_F^T$$

where $\sigma_{\psi}^2 = E[\psi(k)\psi(k)^T]$ is known covariance matrix. Thus the solution to the above Lyapunov equation will give us the desired state and state error co-variance matrices. Substituting the appropriate entries from $\sigma_{y_J}^2$ we get the expected value of the design objective function for the regulatory case. The above derivation gives the expected value of the design cost function for the unconstrained case. This provides an useful lower bound on the performance for the constrained case.

2. **The deterministic case with no model-plant mismatch (open loop case):** Let us consider the case where there is no model plant mismatch and the only external inputs are deterministic changes in the setpoint. In this case the disturbance estimate will reduce to zero since there is no mismatch, $y(k) - \hat{y}(k) = 0$,

$$\hat{J}(k) = \{r_k - F\hat{x}(k) - S\Delta u_k\}\}^T\Gamma\{r_k - F\hat{x}(k) - S\Delta u_k\} + \Delta u_k^T\Lambda\Delta u_k$$
$$\Delta u_k = K(r_k - F\hat{x}(k)).$$

Substituting we get,

$$\hat{J}(k) = (r_k - F\hat{x}(k))^T Q(r_k - F\hat{x}(k))$$
$$E[\hat{J}(k)] = E[(r_k - F\hat{x}(k))^T Q(r_k - F\hat{x}(k))]$$
$$Q = (I - KS)^T\Gamma(I - KS) + K^T\Lambda K$$
$$= \lim_{N\to\infty}\left\{\frac{1}{N}\sum_{k=1}^{N}||\psi(k)||_Q^2\right\}.$$

Finding $E[\hat{J}(k)]$ in this case requires a different approach

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k)$$
$$\Delta u(k) = [I\ 0...0]K(r_k - F\hat{x}(k)) = K_1 r(k) - K_2\hat{x}(k)$$

once again introducing the additional state to take care of the integrator we have,

$$z(k+1) = z(k) + \Delta u(k) = z(k) + K_1 r(k) - K_2\hat{x}(k)$$
$$u(k) = z(k).$$

The state space can be described by

$$\begin{bmatrix} \hat{x}(k+1) \\ z(k+1) \end{bmatrix} = \begin{bmatrix} A & B \\ -K_2 & I \end{bmatrix}\begin{bmatrix} \hat{x}(k) \\ z(k) \end{bmatrix} + \begin{bmatrix} 0 \\ K_1 \end{bmatrix} r(k)$$

or

$$\begin{aligned}\xi(k+1) &= A_F\xi(k) + B_F r(k)\\ \psi(k) &= r_k - F\hat{x}(k)\\ &= \begin{bmatrix} -F & 0 \end{bmatrix}\xi(k) + [I \ldots I]^T r(k)\\ &= C_F\xi(k) + D_F r(k).\end{aligned}$$

We are interested in finding out the weighted root mean squared(RMS)[4] norm of $\psi(k)$ which is equal to the expected value of the design objective function for this case,

$$E(\hat{J}(k)) = \lim_{N\to\infty}\left\{\frac{1}{N}\sum_{k=1}^{N}||\psi(k)||_Q^2\right\}. \tag{5.10}$$

The 2-norm of the above closed loop system will give the objective function for impulse change in the setpoint.

3. **The deterministic case with model-plant mismatch:** Here we consider the case where the true plant is different from the model. This analysis is useful in establishing the impact of modelling uncertainty on the design performance. Let the true plant is given by

$$\begin{aligned}x(k+1) &= Ax(k) + Bu(k)\\ y(k) &= Cx(k)\end{aligned}$$

and the estimator model is given by

$$\begin{aligned}\hat{x}(k+1) &= A'\hat{x}(k) + B'u(k) + P(y(k) - \hat{y}(k))\\ \hat{y}(k) &= C'\hat{x}(k).\end{aligned}$$

The relevant terms in the objective function are given by

$$\begin{aligned}d_k &= [I...I]^T(y(k) - \hat{y}(k))\\ &= [I...I]^T[Cx(k) - C'\hat{x}(k)]\end{aligned}$$

[4]The RMS norm of a signal is defined as: $rms(y) = \lim_{N\to\infty}\sqrt{\frac{1}{N}\sum_{k=1}^{N}||y(k)||^2}$.

and the feedback control law is given by

$$\begin{aligned}\Delta u_k &= K(r_k - F\hat{x}(k) - d_k)\\ &= K(r_k - y_k - (F - [I \dots I]^T C')\hat{x}(k))\\ &= K_1 e(k) - K_2 \hat{x}(k)\\ &= K_1 r(k) - K_1 C x(k) - K_2 \hat{x}(k) - K_1 C v(k)\end{aligned}$$

where $K_1 \triangleq K[I \dots I]^T, K_2 = (F - [I \dots I]^T C')$.

The objective function is described by

$$\begin{aligned}r_k - F\hat{x}(k) - S\Delta u_k - d_k &= r_k - F\hat{x}(k) - S(K_1 e(k) - K_2\hat{x}(k)) - C_1(x(k) - \hat{x}(k))\\ &= ([I \dots I]^T - SK_1)r(k) + (-F + SK_2 + C_1)\hat{x}(k)\\ &\quad -(SK_1C + C_1)x(k)\\ \phi(k) &\triangleq ([I \dots I]^T - SK_1)r(k) + (-F + SK_2 + C_1)\hat{x}(k)\\ &\quad -(SK_1C + C_1)x(k)\\ \Delta u_k &= K_1 r(k) - K_1 C x(k) - K_2 \hat{x}(k)\\ \theta(k) &\triangleq -K_1 x(k) - K_2 \hat{x}(k) - K_1 v(k)\\ \hat{J}(k) &= \phi(k)^T \Gamma \phi(k) + \theta(k)^T \Lambda \theta(k).\end{aligned}$$

The expected value of the objective function can be easily expressed in terms of the powers of the pseudo-outputs defined above,

$$\begin{aligned}E[\hat{J}(k)] &= E[\phi(k)^T \Gamma \phi(k) + \theta(k)^T \Lambda \theta(k)]\\ &= \lim_{N\to\infty} \left\{ \frac{1}{N} \sum_{k=1}^{N} (||\phi(k)||_\Gamma^2 + ||\theta(k)||_\Lambda^2 \right\}.\end{aligned}$$

The state space system can be described by

$$\begin{bmatrix} x(k+1) \\ \hat{x}(k+1) \\ z(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 & B \\ PC & A' - PC' & B' \\ -K_1 C & -K_2 & I \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \\ z(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ K_1 \end{bmatrix} r(k)$$

$$\begin{bmatrix} \phi(k) \\ \theta(k) \end{bmatrix} = \begin{bmatrix} -(GK_1C + C_1) & (-F + GK_2 + C_1) & 0 \\ -K_1 & -K_2 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \\ z(k) \end{bmatrix} + \begin{bmatrix} (I_P - GK_1) \\ 0 \end{bmatrix} r(k)$$

or

$$\begin{aligned}\xi(k+1) &= A_F \xi(k) + B_F r(k) \qquad (5.11)\\ y_J(k) &= C_F \xi(k) + D_F r(k).\end{aligned}$$

For a given $r(k)$ we can find out the power of the pseudo-outputs defined above and they become the value of the achieved objective function for a specific mismatch. The above analysis could be used in conjunction with the multi-model description of uncertainty to evaluate its worst-case effect on the design performance.

4. **The case of deterministic and stochastic inputs:** Let us consider the case where there are both stochastic and deterministic inputs in the system as is the case with most real systems. The true plant is given by

$$x(k+1) = Ax(k) + Bu(k) + B_1 w(k)$$
$$y(k) = Cx(k) + v(k)$$

and the estimator model is given by

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + P(y(k) - \hat{y}(k))$$
$$\hat{y}(k) = C\hat{x}(k).$$

The relevant terms in the objective function are given by

$$d_k = [I...I]^T(y(k) - \hat{y}(k))$$
$$= [I...I]^T[Cx(k) + v(k) - C\hat{x}(k)]$$
$$= I_p Cx(k) + I_p v(k) - I_p C\hat{x}(k)$$

and the feedback control law is given by

$$\Delta u_k = K(r_k - F\hat{x}(k) - d_k)$$
$$= K_1 r(k) - K_1 Cx(k) - K_2\hat{x}(k) - K_1 v(k)$$
$$\text{where } K_1 = KI_P, K_2 = (F - I_P C).$$

The objective function is described by

$$\begin{aligned}
r_k - F\hat{x}(k) - S\Delta u_k - d_k = r_k - F\hat{x}(k) - S(K_1 r(k) - K_1 Cx(k) - K_2\hat{x}(k) \\
-K_1 v(k)) - I_P(Cx(k) + v(k) - C\hat{x}(k)) \\
= (I_p - SK_1)r(k) + (-F + SK_2 + C_1)\hat{x}(k) + \\
(SK_1C - C_1)x(k) + (SK_1 - I_P)v(k) \\
\phi_1(k) \triangleq (I_p - SK_1)r(k) + (-F + SK_2 + C_1)\hat{x}(k) \\
+(SK_1C - C_1)x(k) \\
\phi_2(k) \triangleq (SK_1 - I_p)v(k) \\
\Delta U = K_1 r(k) - K_1 Cx(k) - K_2\hat{x}(k) - K_1 v(k) \\
\theta_1(k) \triangleq K_1 r(k) - K_1 x(k) - K_2\hat{x}(k) \\
\theta_2(k) \triangleq -K_1 v(k) \\
\hat{J}(k) = \phi_1(k)^T \Gamma \phi_1(k) + \theta_1(k)^T \Lambda \theta_1(k) + 2\phi_1(k)^T \Gamma \phi_2(k) \\
+2\theta_1(k)^T \Gamma \theta_2(k) + \phi_2(k)^T \Gamma \phi_2(k) + \theta_1(k)^T \Lambda \theta_2(k).
\end{aligned}$$

The expected value of the objective function can be easily expressed as in terms of the powers of the pseudo-outputs defined above,

$$\begin{aligned}
E[\hat{J}(k)] &= E[\phi(k)^T \Gamma \phi(k) + \theta(k)^T \Lambda \theta(k)] + E[\phi_2(k)^T \Gamma \phi_2(k) + \theta_1(k)^T \Lambda \theta_2(k)] \\
&= \lim_{N\to\infty} \left\{ \frac{1}{N} \sum_{k=1}^{N} (||\phi_1(k)||_\Gamma^2 + ||\theta_1(k)||_\Lambda^2 \right\} + tr\{R\sigma_v^2\}.
\end{aligned}$$

Assuming $[x(k), \hat{x}(k), y_{sp}(k)]$ and $v(k)$ are uncorrelated and

$$R = (SK_1 - I_P)^T \Gamma (SK_1 - I_P) + K_1^T \Lambda K_1$$

the state space system can be described by

$$\begin{bmatrix} x(k+1) \\ \hat{x}(k+1) \\ z(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 & B \\ PC & A - PC & B \\ -K_1 C & -K_2 & I \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \\ z(k) \end{bmatrix} + \begin{bmatrix} 0 & B_1 & 0 \\ 0 & 0 & P \\ K_1 & 0 & -K_1' \end{bmatrix} \begin{bmatrix} r(k) \\ w(k) \\ v(k) \end{bmatrix}$$

$$\begin{bmatrix} \phi(k) \\ \theta(k) \end{bmatrix} = \begin{bmatrix} -(SK_1C + C_1) & (-F + SK_2 + C_1) & 0 \\ -K_1 & -K_2 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \\ z(k) \end{bmatrix}$$

$$+ \begin{bmatrix} (I_P - SK_1) & 0 & -SK_1 \\ 0 & 0 & -K_1 \end{bmatrix} \begin{bmatrix} r(k) \\ w(k) \\ v(k) \end{bmatrix}$$

or

$$\xi(k+1) = A_F\xi(k) + B_F\Psi(k)$$
$$y_J(k) = C_F\xi(k) + D_F\Psi(k).$$

By principle of superposition of linear systems we first calculate the variance contribution to the pseudo outputs defined above from the stochastic inputs assuming the deterministic part or $r(k) = 0$,

$$\phi_1(k) = \phi_1^d(k) + \phi_1^s(k)$$
$$E[\phi_1(k)^T\Gamma\phi_1(k)] = E[\phi_1^d(k)^T\Gamma\phi_1^d(k)] + 2E[\phi_1^d(k)^T\Gamma\phi_1^s(k)] + E[\phi_1^s(k)^T\Gamma\phi_1^s(k)]$$
$$= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}(||\phi_1^d(k)||_\Gamma^2 + tr\{\Gamma E[\phi_1^s(k)\phi_1^s(k)^T\}.$$

Assuming $\phi_1^d(k)$ and $\phi_1^s(k)$ are uncorrelated and $E[\phi_1^s(k)] = 0$. Similarly we can write down for the other term in the objective function,

$$E[\theta_1(k)^T\Gamma\theta_1(k)] = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}(||\theta_1^d(k)||_\Gamma^2 + tr\{\Gamma E[\theta_1^s(k)\theta_1^s(k)^T\}. \tag{5.12}$$

The variance terms and the RMS norm terms can be estimated from the closed loop system defined above using linearity and principle of superposition. Thus the effect of the stochastic and the deterministic components on the cost function can be evaluated separately.

5. **Loss in performance due to the receding horizon implementation.** The receding horizon feature is a unique characteristic of MPC. In the following discussion we determine the effect of the receding horizon nature on design performance. Under the assumptions of closed loop stability we will derive the $\lim_{k\to\infty}\alpha(k)$ for the case of no model plant mismatch,

$$\lim_{k\to\infty}\alpha(k) = \frac{E[\hat{J}(k)]}{E[J_{rec}(k)]}$$

where $J_{rec}(k)$ is obtained by substituting $\Delta u_k^r = [\Delta u(k)\ 0....0]^T$ *i.e.* the receding

horizon implementation instead of the optimal moves,

$$J_{rec}(k) = \{r_k - F\hat{x}(k) - S\Delta u_k^r - d_k\}^T\Gamma\{r_k - F\hat{x}(k) - S\Delta u_k^r - d_k\} + \Delta u_k^{rT}\Lambda\Delta u_k^r$$
$$= \{r_k - F\hat{x}(k) - d_k\}^T\Gamma\{r_k - F\hat{x}(k) - d_k\} + \Delta u_k^{rT}(S^T\Gamma S + \Lambda)\Delta u_k^r$$
$$-2\Delta u_k^{rT}S^T\Gamma\{r_k - F\hat{x}(k) - d_k\}$$
$$= \{r_k - F\hat{x}(k) - d_k\}^T\Gamma\{r_k - F\hat{x}(k) - d_k\} + \Delta u(k)^2 H_{11}$$
$$-2\Delta u(k)J_1^T\{r_k - F\hat{x}(k) - d_k\}$$
$$\text{where } H = (S^T\Gamma S + \Lambda), J = S^T\Gamma \text{ and } J_1^T \text{ is the first row of } J_1$$
$$\Delta u(k) = [I\ 0....0](S^T\Gamma S + \Lambda)^{-1}S^T\Gamma(r_k - F\hat{x}(k) - d_k) = K_1(r_k - F\hat{x}(k) - d_k)$$
$$\text{where } K = (S^T\Gamma S + \Lambda)^{-1}S^T\Gamma \text{ and } K_1 \text{ is the first } m \text{ rows of } K.$$

Substituting for $\Delta u(k)$ we get,

$$J_{rec}(k) = \{r_k - F\hat{x}(k) - d_k\}^T(\Gamma + Q_1 - Q_2)\{r_k - F\hat{x}(k) - d_k\}$$
$$\text{where } Q_1 = K_1^T H_{11} K_1, Q_2 = 2K_1^T J_1^T.$$

If we compare the receding horizon objective function to the design objective function we have

$$\hat{J}(k) = \{r_k - F\hat{x}(k) - S\Delta u_k - d_k\}^T\Gamma\{r_k - F\hat{x}(k) - S\Delta u_k - d_k\} + \Delta u_k^T\Lambda\Delta u_k$$
$$= \{r_k - F\hat{x}(k) - d_k\}^T\Gamma\{r_k - F\hat{x}(k) - d_k\} + \Delta u_k^T(S^T\Gamma S + \Lambda)\Delta u_k$$
$$-2\Delta u_k^T S^T\Gamma\{r_k - F\hat{x}(k) - d_k\}$$
$$\text{Since } \Delta u_k = (S^T\Gamma S + \Lambda)^{-1}S^T\Gamma\{r_k - F\hat{x}(k) - d_k\} \text{ we have}$$
$$\hat{J}(k) = \{r_k - F\hat{x}(k) - d_k\}^T\Gamma\{r_k - F\hat{x}(k) - d_k\} - \Delta U^T S^T\Gamma\{r_k - F\hat{x}(k) - d_k\}$$
$$= \{r_k - F\hat{x}(k) - d_k\}^T(\Gamma - Q)\{r_k - F\hat{x}(k) - d_k\}$$
$$\text{where } Q = \Gamma^T S(S^T\Gamma S + \Lambda)S^T\Gamma.$$

If we define

$$\phi(k) = \{r_k - F\hat{x}(k) - d_k\}$$

then we have

$$E[\hat{J}(k)] = tr\{(\Gamma - Q)\sigma_\phi^2\} \tag{5.13}$$
$$E[J_{rec}(k)] = tr\{(\Gamma - Q_1 - Q_2)\sigma_\phi^2\}.$$

Therefore

$$\lim_{k\to\infty} \alpha(k) = \frac{tr\{(\Gamma - Q)\sigma_\phi^2\}}{tr\{(\Gamma - Q_1 - Q_2)\sigma_\phi^2\}} \tag{5.14}$$

$$= \frac{tr\{(\Gamma - Q)\sigma_\phi^2\}}{tr\{(\Gamma - Q)\sigma_\phi^2\} + tr\{(Q - Q_1 - Q_2)\sigma_\phi^2\}}$$

$$= \frac{1}{1 + tr\{(Q - Q_1 - Q_2)\sigma_\phi^2\}/tr\{(\Gamma - Q)\sigma_\phi^2\}}.$$

It is obvious that $(Q - Q_1 - Q_2)$ is positive semidefinite (since $J_{rec}(k) \geq \hat{J}(k)$). σ_ϕ^2 can be found by the method described earlier. The above expression quantifies the loss in performance due to the receding horizon nature of the control law. If $m = 1$, $Q = Q_1 + Q_2$ and there is no loss in performance due to the receding horizon nature of the control law. Subsequent simulation examples illustrate the loss in performance due to receding horizon implementation.

6. **Convergence of the performance index, $\alpha(k)$** : It is important to know that the performance index, $\alpha(k)$, will not diverge provided the closed loop system is stable. What conditions are necessary to prove the convergence of the performance index? Let us consider the case where we are finding the achieved objective function from the true data,

$$J(k) = \{r_k - y_k\}^T \Gamma \{r_k - y_k\} + \Delta u_k^T \Lambda \Delta u_k$$
$$= ||r_k - y_k||_\Gamma^2 + ||\Delta u_k||_\Lambda^2.$$

Therefore we have

$$J(k) = ||r_k - \hat{y}_k + \hat{y}_k - y_k||_\Gamma^2 + ||\Delta u_k^* - \Delta u_k^* + \Delta u_k||_\Lambda^2$$
$$\leq ||r_k - \hat{y}_k||_\Gamma^2 + ||\hat{y}_k - y_k||_\Gamma^2 + ||\Delta u_k^*||_\Lambda^2 + ||\Delta u_k - \Delta u_k^*||_\Lambda^2$$
$$= \hat{J}(k) + ||\hat{y}_k - y_k||_\Gamma^2 + ||\Delta u_k - \Delta u_k^*||_\Lambda^2$$
$$J(k) \geq |\hat{J}(k) - ||\hat{y}_k - y_k||_\Gamma^2 - ||\Delta u_k - \Delta u_k^*||_\Lambda^2|$$
$$J(k) = \hat{J}(k) + (\hat{y}_k - y_k)^T \Gamma (\hat{y}_k - y_k) + (\Delta u_k - \Delta u_k^*)^T \Lambda (\Delta u_k - \Delta u_k^*)$$
$$-2(r_k - \hat{y}_k)^T \Gamma (\hat{y}_k - y_k) - 2\Delta u_k^{*T} \Lambda (\Delta u_k^* - \Delta u_k).$$

Thus the achieved objective function equals the design objective function plus two terms which are positive and additional two terms which could be either positive or negative. Whether the achieved performance betters the design performance depends

on how these last two terms behave,

$$\frac{1}{1+J_e(k)/\hat{J}(k)} \leq \frac{\hat{J}(k)}{J(k)} \leq \frac{1}{|1-J_e(k)/\hat{J}(k)|}$$
$$J_e(k) = ||\hat{y}_k - y_k||^2_\Gamma + ||\Delta u_k - \Delta u^*_k||^2_\Lambda.$$

In the limiting case if the upper bound on $\alpha(k)$ is finite and bounded then it implies that $\lim_{k\to\infty} \alpha(k)$ if it exists, is finite, *i.e.* the series $\alpha(k)$ is not divergent. We assume that for a closed loop stable system,

$$||\hat{y}_k - y_k||^2_\Gamma \leq \bar{\gamma}, ||\Delta u_k - \Delta u^*_k||^2_\Lambda \leq \bar{\rho}, \text{for all } k \geq 0.$$

Therefore

$$E[J(k)] \leq E[\hat{J}(k)] + \bar{\gamma} + \bar{\rho}$$
$$E[J(k)] \geq E[|\hat{J}(k) - \bar{\gamma} - \bar{\rho}|] \geq |E[\hat{J}(k)] - \bar{\gamma} - \bar{\rho}|$$
$$\text{If } E[\hat{J}(k)] \geq \bar{\gamma} + \bar{\rho}.$$

Hence

$$\frac{1}{1+(\bar{\gamma}+\bar{\rho})/E[\hat{J}(k)]} \leq \lim_{k\to\infty} \alpha(k) = \frac{E[\hat{J}(k)]}{E[J(k)]} \leq \frac{1}{|1-(\bar{\gamma}+\bar{\rho})/E[\hat{J}(k)]|} \tag{5.15}$$

Thus under some assumptions, the performance index will remain bounded. In fact, equation 5.15 provides useful lower and upper bounds on the performance index. The tighter these bounds are the closer the achieved performance will be to the design performance. In the subsequent Sections, simulation examples and an industrial application will demonstrate some of the properties proved above.

5.4 Simulation examples

5.4.1 Mixing Process

The above approach was applied to a simulation example. The system under consideration is a 2x2 mixing process. The controlled variables are temperature (y_1) and water level (y_2) and the manipulated inputs are inlet hot water (u_1) and inlet cold water (u_2) flow rates. The following model is available in discrete form,

$$P(z^{-1}) = \begin{bmatrix} \frac{0.0235z^{-1}}{1-0.8607z^{-1}} & \frac{-0.1602z^{-1}}{1-0.8607z^{-1}} \\ \frac{0.2043z^{-1}}{1-0.9827z^{-1}} & \frac{0.2839z^{-1}}{1-0.9827z^{-1}} \end{bmatrix}.$$

A MPC controller was used to control this process in the presence of unmeasured disturbances. The controller design parameters were:

$$p = 10, m = 2, \Lambda = diag([1, 4]), \Gamma = diag([1, 2]).$$

White noise sequences at the input and output with covariance equal to $0.1I$ served as the unmeasured disturbances. First the LQG benchmark was found, and the performance of a constrained and unconstrained MPC was evaluated against this benchmark (see figure 5.6). The constraints on input moves were artificially imposed in order to activate the constraints frequently. The unconstrained controller showed better performance compared to the constrained controller with respect to the LQG benchmark.

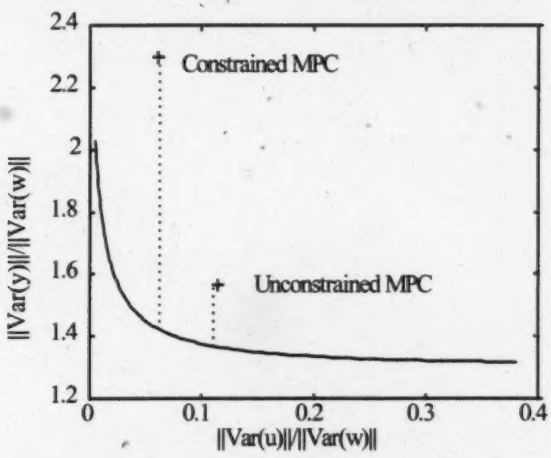


Figure 5.6: *MPC performance assessment using the LQG benchmark*

A plot of the LQG objective function compared to the achieved objective function is shown in figure 5.7. A performance measure of 0.6579 was obtained for the unconstrained controller.

Performance assessment of the same controller using the design case benchmarking approach yields contrasting results (Table 5.1). For the unconstrained controller a performance index 0.8426 revealed satisfactory performance while the imposition of constraints led to a performance index of 1. The constrained controller showed improvement according to one benchmark and deterioration with respect to the LQG benchmark. The design case approach indicates that the controller is doing its best under the given constraints while the LQG approach which is based on comparison with an unconstrained controller shows a fall in performance.

Figure 5.8 shows the input moves during the regulatory run for the constrained controller. The constraints are active for a large portion of the run and are limiting the

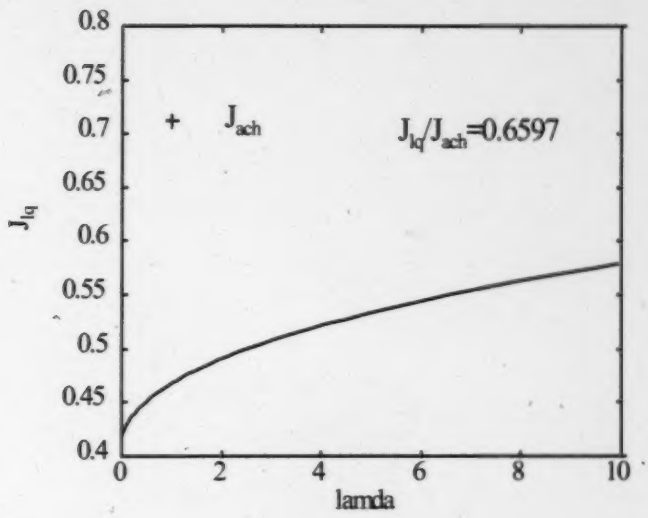


Figure 5.7: *Comparison of the achieved performance with the LQG objective function.*

Table 5.1: *Effect of constraints on MPC performance*

	LQG	$E(\hat{J})/E(J)$
Unconstrained	0.6579	0.8426
Constrained	0.4708	1.00

performance of the controller in an absolute sense (LQG). On the other hand the controller cannot do any better due to design constraints as indicated by the design case benchmark.

5.4.2 Effect of reduced order modeling

A contrived example was considered to show the effect of mismatch on such performance measures. The true plant is a third order overdamped process: $G(s) = 1/(s+1)(3s+1)(5s+1)$ whose discrete equivalent for a sampling time of $T_s = 1$ is

$$G(z^{-1}) = \frac{0.0077z^{-1} + 0.0212z^{-2} + 0.0036z^{-3}}{1 - 1.9031z^{-1} + 1.1514z^{-2} - 0.2158z^{-3}}. \tag{5.16}$$

This *plant* is approximated by the following 1^{st} order model which is estimated through identification techniques:

$$G(z^{-1}) = \frac{0.0419z^{-1} + 0.0719z^{-2}}{1 - 0.8969z^{-1}}. \tag{5.17}$$

In practice such approximations are common and the controller design is often based on reduced order models. It should also be noted here that both the plant and the model have

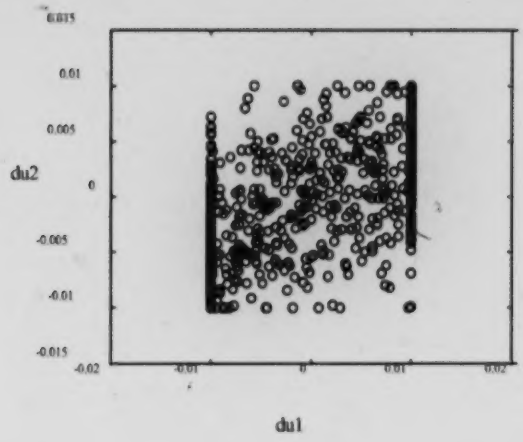


Figure 5.8: *The input moves for the constrained controller during the regulatory run*

non-minimum phase zeros. The presence of non-minimum phase zeros is known to pose fundamental limitations on achievable performance. The nominal regulatory performance of a MPC controller relative to the optimal performance curve is shown in figure 5.9. The effect of the prediction horizon on regulatory performance is illustrated in table 5.2. A reduction in the prediction horizon leads to improved regulatory performance according to both performance measures.

Table 5.3 shows the tracking and regulatory performances of the MPC based on the design case benchmark for (1) the nominal case or the design case, (2) the achieved case with model plant mismatch and (3) the achieved case for the constrained controller. The MPC tuning parameters were:

$$P = 8, M = 2, \Lambda = I, \Gamma = I$$
$$-1 \leq \Delta u(k) \leq 1, -10 \leq u(k) \leq 20.$$

The MPC controller shows acceptable tracking performance but poor regulatory performance. For the regulatory case unmeasured disturbance were added at the input and output with variances equal to 0.01. For the tracking case a square wave of magnitude ± 1 was used. A comparison of rows 1 and 2 of table 5.3 shows clearly the effect of model plant mismatch on performance. Again, the imposition of constraints led to improvement in performance (see rows 2 and 3). A comparison of columns 1 and 2 with column 3 shows that the tracking performance dominates the combined case leading to a satisfactory overall performance. Table 5.4 shows the regulatory performance for stochastic and deterministic

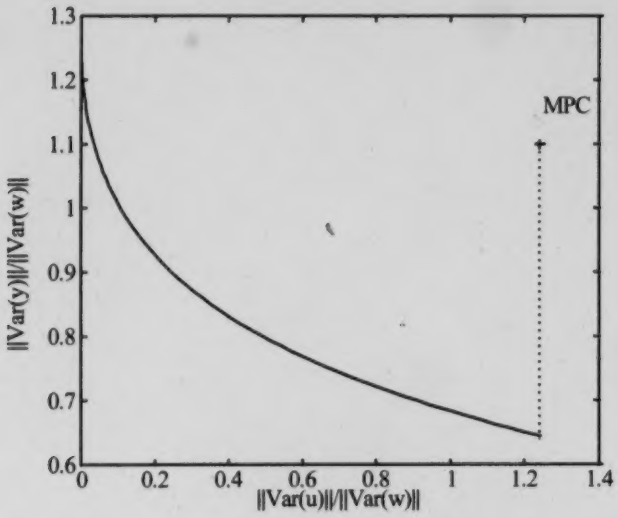


Figure 5.9: *Regulatory performance of a model predictive controller.*

step type disturbances. The stochastic disturbances were added at the input and output with variances equal to 0.01. The deterministic disturbance was a square wave of magnitude ± 0.25 added at the output. The first column of this table is similar to column 2 of table 5.3. The MPC controller shows poor disturbance rejection for either type of disturbance. The model plant mismatch does not affect the regulatory performance significantly. A similar observation can be made in the presence of constraints. The constraints were never activated in the regulatory runs.

Figures 5.10 (a) and (c) show the comparison of the design and achieved objective functions for different control weightings. For $\lambda = 0$ the achieved objective function is significantly higher than the design values and the system shows a poorly damped servo response (figure 5.10 (b)). For $\lambda = 0.5$ the design objective has higher values and is closer to the achieved values. The result is a comparatively satisfactory servo response (figure 5.10 (d)). It should be noted here that as a result of higher λ, the design performance requirements were lowered in order to reduce the difference between the achieved and the design performance (performance degradation). For $\lambda = 0$ the design requirements proved to be too stringent and the achieved performance deteriorated considerably.

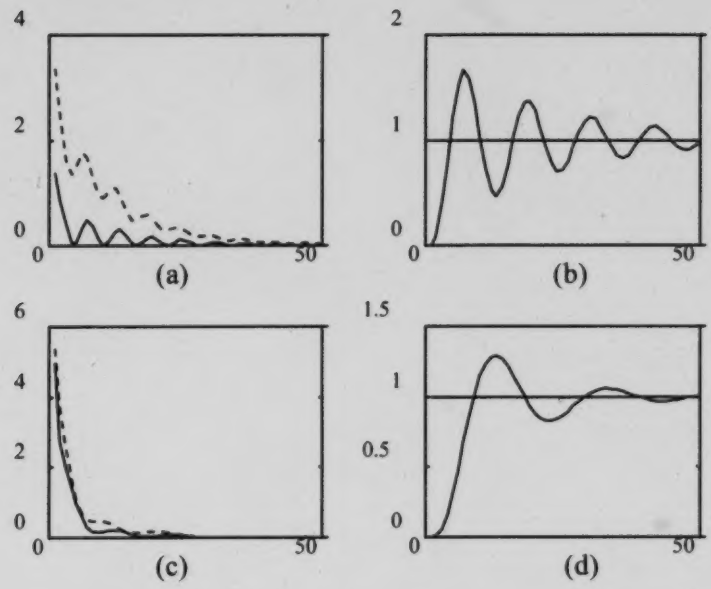


Figure 5.10: *Comparison of design (–) and achieved performance objectives (- -) for* $\lambda = 0$ *(a) and* $\lambda = 0.5$ *(c) and the corresponding servo responses achieved by MPC - (b) and (d)*

5.5 Application to QDMC assessment

A QDMC controller in Shell's hydrocracker unit (HCU) was chosen for assessing QDMC performance. The QDMC manipulates the mild vacuum column (MVC) bleed flow and the two second stage weighted average bed temperatures (WABTs) to control the recycle surge drum level. The real CV in the controller formulation is the change in level (Δh) . Since the application controls one level with two WABTs, an additional controlled variable, the Δ WABT was added. The WABT is the difference between the train 1 and 2 second stage WABTs. Its setpoint may be adjusted to balance the hydrogen consumption gap between the trains. Figure 5.11 shows a schematic of the process and the position of the mild vacuum column with respect to the reactors whose WABTs are the MVs. The QDMC is implemented once every 10 minutes. This QDMC application currently has 2 controlled variables (CVs), 3 manipulated variables (MVs), 1 permanently constrained manipulated variable (PCMV), 3 associated variables (AVs) and 2 feedforward variables (FFs). The prediction horizon $p = 20$ and the control horizon $m = 5$.The HCU is in material balance whenever the recycle drum level is steady, indicating the reaction severity is just enough to crack all recycle material to extinction (Shell Application Manual - Howie, 1995). This

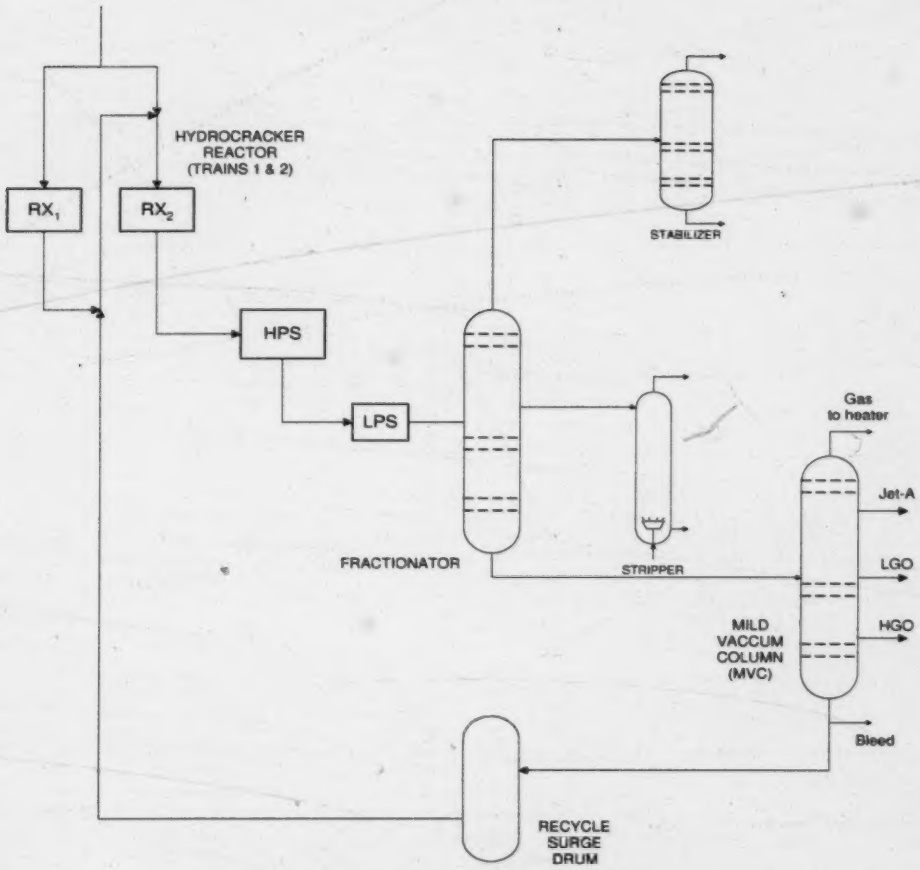


Figure 5.11: *A schematic overview of the process*

condition is necessary to properly optimize the unit, whether through manual adjustments to such variables as CFR or product cutpoints, or through on-line moves made by the HCU optimizer. The purpose of this application is to make adjustments to reaction severity in order to bring the unit back into balance after a disturbance has occurred.

The bleed flow rate has been included in this control application as a MV so that short term fluctuations in the level will be carried out of the unit by adjusting the bleed rather than being kept within the unit. The application will adjust the second stage WABTs slowly over a longer time frame to keep the daily average bleed flow at the desired target.

As long as the drum level is steady and remains within the constraint limits, the controller will make minimal moves to the second stage WABT setpoints of both trains and the bleed flow rate. Larger moves will be made when the level changes quickly and when it appears that the recycle drum level high or low limit will be violated. The difference between the reactor severity in each of the HCU trains is maintained by controlling the Δ WABT: the difference between train 1 and train 2 second stage WABTs.

Total fresh HCU feed flow and the HGO draw flow are included as feedforward variables in the QDMC application currently. The controller will anticipate the effect of changes in these variables when computing the size and direction of moves to be made to reactor WABT's. Included amongst the associated variables are:

1. Recycle drum level with the purpose of providing a means of level flow smoothing i.e., the level is allowed to float within this range with fewer moves by the controller, but once outside this range the control moves are more aggressive. The level is constrained to lie within 40 and 60% of the maximum.

2. To prevent the MVC filter temperature limit trip, the temperature is constrained by a maximum limit.

3. To maintain MVC bottoms level, the valve position of the level controller is constrained to lie above 5%. The level could otherwise be easily lost due to excessive bleed flows.

The degree to which the constraint limits are followed and the controlled setpoints are relaxed depends on how the application tuned. The QDMC application is also programmed to deal with hard constraints on the manipulated variables and their rate of change.

Figure 5.12 shows the step response models between the MVs,FFs and the CVs, AVs. A QDMC based on the step response models is used to control the surge drum level.3

Table 5.2: *Effect of prediction horizon on regulatory performance*

	LQG	Design vs. Achieved
P=8	0.5161	0.6875
P=5	0.7027	0.8564

Table 5.3: *Tracking and regulatory performance*

	Tracking	Regulatory	Combined
Nominal	0.9036	0.4724	0.8441
Achieved	0.7547	0.4621	0.7474
Constrained	0.8120	0.4621	0.8038

Table 5.4: *Disturbance rejection properties of MPC*

	Stochastic	Deterministic	Combined
Nominal	0.4721	0.4295	0.4559
Achieved	0.4617	0.4083	0.4353
Constrained	0.4617	0.4087	0.4353

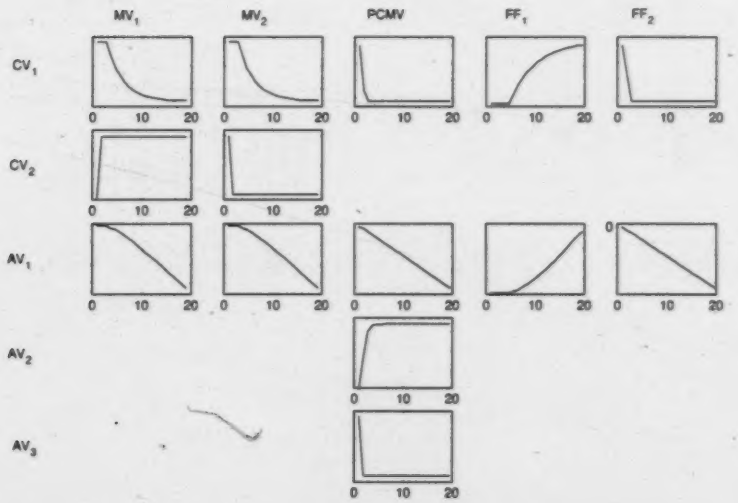


Figure 5.12: *The step response models between the active inputs and outputs.*

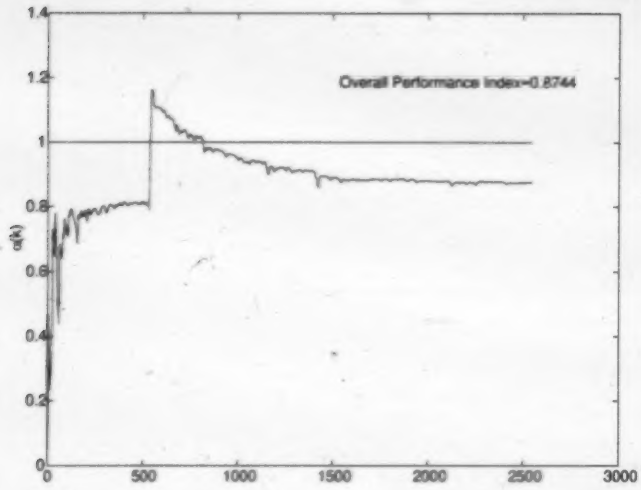


Figure 5.13: *Performance Measure based on comparison of the design and achieved objective functions.*

Data was collected for over two weeks and the next Section deals with the performance of the QDMC based on this data.

Design vs. Achieved Objective function: The on-line predicted values were available for the CVs, AVs and MVs. These projections were used to calculate the design objective function and this was compared with achieved objective function. Figure 5.13 shows the ratio of the cumulative design and achieved objective functions, $\alpha(k)$. The overall achieved performance is 87.44% of the design performance amounting to an average 12.56% degradation of performance. The sudden shift in the performance index was found to correspond with the large drop in the total fresh feed flow rate (see figure 5.14).

Figure 5.15 shows the mean values of the design and the achieved objective functions along with their components as discussed in Section 5.3. Once again the large increase in the design objective function coincides with the drop in fresh feed. This observation leads us to conclude that the increase in performance index in figure 5.13 is due to a lowering of the design performance whereas the achieved performance is relatively unchanged. The sudden drop in FF1 is later confirmed where the flow meter was taken out of service for recalibration. The actual fresh feed is unchanged but a false reading is used in QDMC to produce unnecessary moves. For the achieved objective function, the CV (78.43%) and the MV (15.59%) terms are the two major contributors, whereas for the design objective function the CV (84.42%) and the AV (15.44%) terms are dominant. Table 5.5 shows the

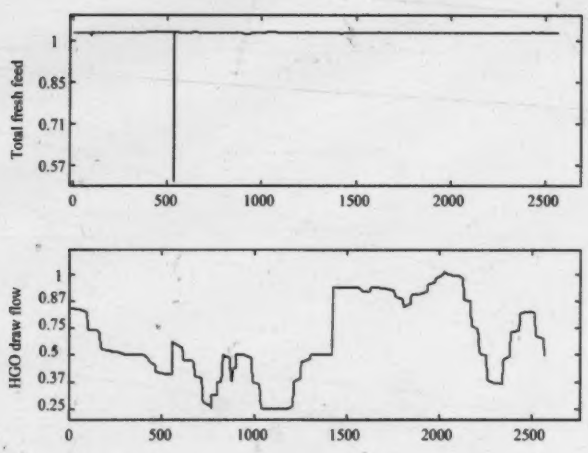


Figure 5.14: *The FFs time series plot.*

Table 5.5: *Contribution of various terms to the design and achieved objective functions*

$E(\hat{J})$	CV	MV	PCMV	AV
5.22	84.42%	0.08%	0.05%	15.44%
$E(J)$	CV	MV	PCMV	AV
5.96	78.43%	15.59%	1.31%	4.68%

final mean value of the objective function and the percentage contributions of the various terms.

Figure 5.16 shows the performance indices based on the individual terms in the objective function. For the CVs the achieved performance is almost equal to the design performance (94.13%), the MVs and the PCMVs the achieved is much worse than the designed performance at 0.47% and 3% respectively. For the AVs alone, the achieved performance is much better than design performance (288.4%).

To summarize the outcome of this industrial case study:

- The QDMC application for the recycle drum level control is delivering satisfactory performance. The performance index is 0.87, according to the objective function method.

- The constraint handling, however, is not satisfactory. Constraints on the drum level are violated 24.5% of the time. Constraint violations also cause poor performance in

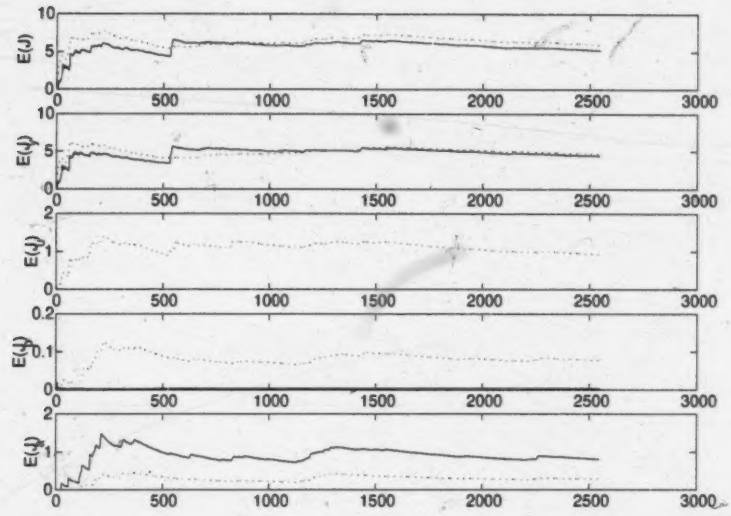


Figure 5.15: *The mean values of the design (solid line) and the achieved objective functions (dotted line) shown along with their components - J_y, J_u, J_p and J_a.*

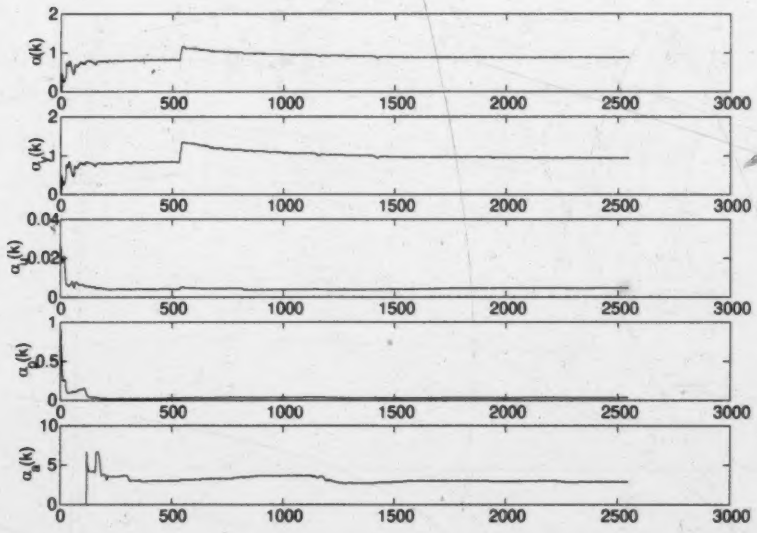


Figure 5.16: *Performance indices based on comparison of the different components in the objective function.*

CV2, the difference between the WABT setpoints for train 1 and train 2. The possible reasons include a poor model for AV1 and/or unmeasured disturbances.

- There appears to be a sensor false reading for FF1, the total fresh feed flow at sampling instant 533. This was confirmed by noticing that the achieved objective function is unchanged whereas the design objective function undergoes a large change at this point in time. The prediction error analysis also led to a similar conclusion since the model predictions showed a large upset and the measured values of the outputs were relatively unchanged.

5.6 Conclusions

Performance assessment of model predictive controllers based on (1) the LQG benchmark and (2) the design case benchmark, is recommended in this chapter. The use of the LQG benchmark requires knowledge of the process and noise dynamics. In practice only approximate models are available for either the process or disturbances. This can lead to inaccurate benchmarking via the LQG approach. Our future goals include the incorporation of the effect of modeling uncertainty into the LQG benchmarking scheme.

The design case benchmark proposed in this work can be treated as a relative measure of performance. The nominal or the design case serves as the basis for comparison with the achieved performance. This approach makes indirect use of the prediction model used by the controller. The multivariable nature of most chemical processes and presence of nonlinearities does not pose a hindrance to the application of this method. The proposed measure of performance has been shown to be sensitive to model plant mismatch, hard constraints, stochastic and deterministic disturbances through two simulation examples. Some theoretical properties of this index were established. These include analytical estimation of the design objective function, effect of the receding horizon policy on MPC performance and conditions for convergence of the proposed index. An industrial case study was used to illustrate the usefulness of this performance metric.

In summary the two approaches presented here are two different ways of evaluating performance, the LQG approach is an absolute measure of performance while the design case benchmark is a relative measure. The next logical step in performance assessment is to determine the causes of poor performance as defined by either of the above measures.

Chapter 6

Issues in Performance Diagnostics of Model-based Controllers

6.1 Introduction

One of the primary objectives of performance assessment is to evaluate existing controllers using routine operating data. Several researchers have proposed ways of automating this assessment process through time series analysis and comparison with the minimum variance controller (Harris, 1989; Huang *et al.* 1997). A poor performance index can be the outcome of one or more of the following: (1) Poor controller design (2) Varying disturbances (3) Limits on achievable performance with existing process control structure (4) Design constraints (5) Changes in plant dynamics. After a preliminary investigation if the control engineer finds that the controller is not doing well, one would also like to have a diagnosis *i.e.* establish the reasons for poor performance.

Most of the recent work in performance assessment has focused on obtaining statistics for performance from routine operating data. Little attention has been paid to the area of diagnostics. If the statistics indicate poor performance there are no systematic ways of detecting the underlying causes. Kesavan and Lee (1997) addressed this issue in the context of multivariable model based controllers. They base their diagnostic scheme on a mechanistic model of the process and analysis of the prediction errors. Presence of closed loop excitation is necessary in order to implement their scheme. Stanfelj *et al.* (1993) use the cross-correlation test to distinguish between plant modelling error and disturbance modelling error. The cross-correlation between the setpoint changes and the model plant mismatch is checked to verify which model needs attention. Tyler and Morari (1996) expressed a large number of performance measures as constraints on the FIR coefficients of

[1] A version of this chapter has been accepted for presentation as "The Effect of Uncertainty on Controller Performance", 1999, R. S. Patwardhan, S. L. Shah and B. Huang, IFAC World Congress, Beijing, China.

a pre-whitening filter. Subsequently hypothesis testing was used to detect if any of these constraints are being violated leading to a deterioration in performance. More recently Gustafsson and Graebe (1998) propose the use of statistical hypothesis testing to detect whether an observed change in nominal performance is due to a disturbance or due to a change in the process that has affected the robustness of the closed loop system.

The objective of controller design is to optimize the performance under feedback. The characteristics of a problem decide the choice of the design objective. Of primary importance is the stability of the closed loop control system. Conventionally a single measure of the tracking/regulatory performance in either the time domain or the frequency domain forms the design objective. More often than not the process has features which limit the achievable performance of any controller. Amongst these limiting features are the presence of time delays, non-minimum phase zeros, constraints on the inputs and outputs, *etc.*

It is important to know at the outset what is the contribution of these features towards the minimum achievable performance. The knowledge of the minimum achievable performance is useful in assessing the performance of any controller. The contribution of the time delay term towards minimum achievable variance is well known and this forms the basis of the minimum variance benchmarking in performance assessment (Huang, 1997). In this chapter we have focused on the *quantifying* effect of (i) non-minimum phase zeros, (ii) input and output constraints and (ii) the presence of modelling uncertainty and disturbance uncertainty and (iv) the impact of nonlinearity on the performance of a linear model-based controller.

The rest of this chapter is organized as follows: Section 6.2 gives the preliminaries required for further analysis. This includes discussion of conventional feedback control, introduction to some common norms for signals and systems and the internal model control framework which we adopt throughout the chapter. Section 6.3 reviews known results on the fundamental limitations on performance of a discrete controller due to (i) discretization, (ii) nonminimum phase zeros and (iii) hard constraints. An interesting results due to Toker *et al.* (1998) points out that directionality of external inputs can significantly affect the system performance. In section 6.4 we derive bounds for the worst case achieved performance for a uncertain system. The effect of uncertainty in time delay is highlighted and the results are extended to the multivariate case. The combined impact of deterministic and stochastic inputs is illustrated. The results are extended to the multivariate case and the effect of uncertainty in the diagonal interactor on multivariable performance measures is established. The impact of nonlinearity on the performance of a linear model based

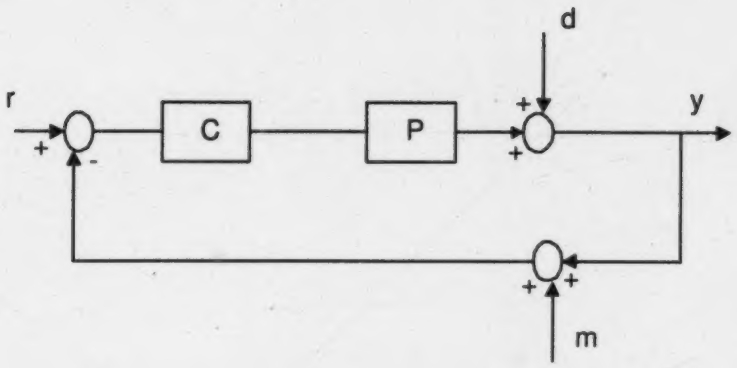


Figure 6.1: *The conventional feedback system*

controller is highlighted in section 6.5. The nonlinearity measure of Allgower (1995) as defined under feedback is utilized in obtaining a bound on the deviation from the design performance. A discussion of model predictive control is included. Simulation examples are used to illustrate the main results in section 6.6. Section 6.7 gives the concluding remarks.

Throughout this chapter we have restricted ourselves to the case of stable plants. We are looking at the class of discrete systems which are finite dimensional linear time invariant (FDLTI). For the sake of brevity, we have omitted the argument z in writing the relations between the signals and the transfer functions for discrete systems. Most of the theory developed here is for SISO systems although it can be easily extended to the multivariable case.

6.2 Preliminaries

Figure 6.1 shows the standard feedback configuration. The measurement error m is assumed to be unknown and most frequently described as a white noise process. The disturbance d is made up of measured and unmeasured disturbances. The measured disturbances may be compensated via a feedforward controller.

For a SISO process, the closed loop relations between the inputs and the outputs can be written as:

$$(1+PC)y = d + PC(r-m).$$

If we define the sensitivity function as

$$S = (1+PC)^{-1}.$$

and the complementary sensitivity function or the closed loop transfer function as

$$T = (1 + PC)^{-1}PC$$

then we have

$$y = Sd + Tr - Tm. \tag{6.1}$$

With this controller structure we have only one degree of freedom (C). Ideally we would like to have S as small as possible at all frequencies to ensure good disturbance rejection and T close to unity for acceptable tracking performance. On the other hand we would like T to be small for the feedback loop to be insensitive to measurement errors. From the above relations we have

$$T + S = 1. \tag{6.2}$$

Hence if we make S zero, then T will be unity and vice-versa. This leads to an unavoidable trade-off in controller design between tracking and measurement noise attenuation. In process control usually the setpoint changes and the disturbances occur in the low frequency region, while the spectrum of measurement errors is uniformly distributed over a wide frequency range. The solution often adopted is to make $|S(e^{jw})|$ small at low frequencies ($|T(e^{jw})| \approx 1$) and $|T(e^{jw})|$ small at high frequencies. Presence of right half plane zeros/poles poses fundamental limitations on the range of frequencies over which sensitivity function can be reduced. For a more detailed discussion of these factors the interested reader is referred to Maciejowski (1989). Skogestad and Postlethwaite (1996) also give an informative discussion on the role of different factors such as RHP zeros/poles, input constraints, uncertainty, exogenous inputs in limiting achievable controller performance for SISO and MIMO systems.

Some idea of the *size* of a signal and the relevant transfer functions, such as the sensitivity function, are needed in order to quantify performance of a control system. Two norms for capturing the magnitude of a signal are:

$$\begin{aligned} \text{2-norm} \quad &: \quad ||u||_2 = \{u(0)^2 + u(1)^2 + ...\}^{1/2} \\ \infty\text{-norm} \quad &: \quad ||u||_\infty = \sup_k |u(k)|. \end{aligned} \tag{6.3}$$

The following system norms provide ways of capturing the magnitude of a linear time

invariant system (Chen and Francis, 1995):

$$\text{1-norm} \quad : \quad ||P||_1 = |P(0)| + |P(1)| + ... = \sup_{||u||_\infty \neq 0} \frac{||y||_\infty}{||u|_\infty} \tag{6.4}$$

$$\text{2-norm} \quad : \quad ||P||_2 = (\frac{1}{2\pi} \int_0^{2\pi} |P(e^{jw})|^2 dw)^{1/2} = P(0)^2 + P(1)^2 + ...$$

$$\infty\text{-norm} \quad : \quad ||P||_\infty = \sup_w |P(e^{jw})| = \sup_{||u||_2 \neq 0} \frac{||y||_2}{||u||_2}.$$

The 1-norm is defined in terms of the impulse response -$P(0), P(1)$...., whereas the 2-norm and the ∞-norm are defined in terms of the transfer functions. The 2-norm of a signal is invariant with respect to the Z-transform. The ∞-norm obeys the following property, in addition to basic properties that define a norm:

$$||P_1 P_2||_\infty \leq ||P_1||_\infty ||P_2||_\infty. \tag{6.5}$$

In fact this property is satisfied by any induced norm such as the 1-norm. A detailed discussion on signal and system norms and their properties, for univariate as well as multivariate systems, can be found in Boyd and Barrat (1991).

The internal model control (IMC) framework provides a way of re-paramaterizing the conventional controller (Morari and Zafiriou, 1989). It is well known that this approach is equivalent to the celebrated Youla-Kucera parameterization of all stabilizing controllers for a stable system. Figure 6.2 shows a schematic of the IMC framework. The controller Q in the IMC framework is related to the conventional feedback controller C in figure 1 by

$$Q = C(I + \hat{P}C)^{-1}.$$

For stable plants a stable Q guarantees internal stability. The nominal sensitivity function and the complimentary sensitivity function can be expressed in terms of Q as

$$\hat{S} = (I - Q\hat{P}), \hat{T} = Q\hat{P}. \tag{6.6}$$

In the remaining part of this chapter we use the IMC framework for analysis purposes.

6.3 Fundamental Limitations on Performance

6.3.1 Effects of discretization

One of the most fundamental limitations on the performance of a discrete controller is due to the sampling process. The sampling rate determines the information available to the discrete compensator and limits its performance (Morari and Zafiriou, 1989; Chen and Francis,

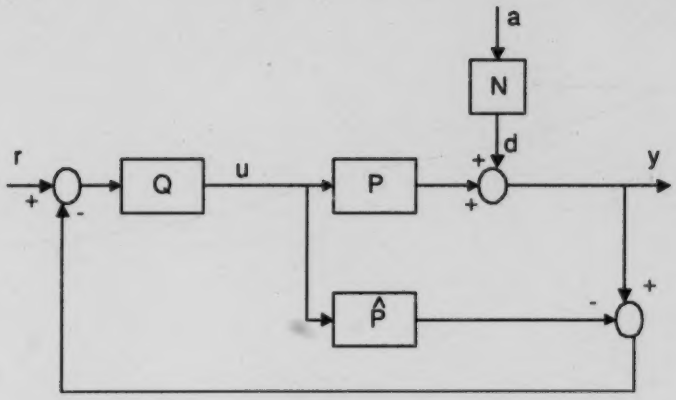


Figure 6.2: *The internal model control stucture*

1995). An inappropriate choice of the sampling interval can lead to loss of controllability for the discrete systems (Chen and Francis, 1995). A large sampling interval relative to the dominant process time constant limits the range of frequencies over which disturbances can be attenuated. A poorly designed discrete time regulator can lead oscillatory intersample behaviour. These and other effects of discretization are well studied and we refer the interested reader to Morari and Zafiriou (1989) and Chen and Francis (1995) for a detailed discussion of these issues.

6.3.2 Limitations due to non-minimum phase zeros

Non-minimum phase zeros pose limitations on achievable performance under feedback. It is a well known fact in performance assessment that the non-minimum phase part, imposes the most basic limitations on minimum achievable variance. The minimum variance benchmark is based on the idea that first few impulse response coefficients of the closed loop system are controller invariant. In general we can have non-minimum phase zeros. The frequency domain effects of such elements is well known, non-minimum phase zeros pose limitations on the shaping of the sensitivity function. The following result from Morari and Zafiriou (1989) is useful in quantifying the effect of non-minimum phase zeros on achievable controller performance.

Theorem 6.1. *Assume that $\hat{P}$ is stable. Factor the model, $\hat{P}$ into an all pass part $\hat{P}_A$ and a minimum phase part $\hat{P}_M$.*

$$\hat{P} = \hat{P}_A \hat{P}_M$$

where

$$\hat{P}_A = z^{-N} \prod_{j=1}^{h} \frac{(1-(a_j^*)^{-1})(z-a_j)}{(1-a_j)(z-(a_j^*)^{-1})}$$

and $a_j, j = 1,..,h$ are the zeros of P which are outside the unit disk (D). The positive integer N, is chosen such that P_M is proper i.e. its numerator and denominator have the same degree. Factor the input, $\hat{v}(z)$ similarly - i.e.

$$\hat{v}(z) = \hat{v}_A(z)\hat{v}_M(z)$$

The H_2 optimal controller is given by

$$Q_{H_2}(z) = z(\hat{P}_M v_M)^{-1}\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_*$$

where the operator $\{.\}_*$ denotes that after a partial fraction expansion of the operand only the strictly proper and stable (including poles at z=1) terms are retained.

The proof for the general case can be found in Morari and Zafirou (1989). Here we prove this theorem for the special case where the input ν does not have poles on the unit circle. Before embarking on the proof we need to briefly introduce the concept of H_2 spaces (Chen and Francis, 1995). All real-rational, proper transfer functions with no poles in the inside or on the unit disk make up the H_2 space. Its complement, the $H_2^\perp$ space contains those transfer function having no poles outside or on the unit disk. These two spaces are orthogonal complements as defined by the inner product of two transfer functions[2]. For SISO systems this decomposition can be obtained by a partial fraction expansion.

[2]The inner product of two transfer functions is defined by $\langle f,g \rangle \triangleq \frac{1}{2\pi}\int_0^{2\pi} f^*(e^{jw})g(e^{jw})dw$ where the operator * denotes the complex conjugate.

The above result is obtained by minimizing the 2-norm of the error:

$$\begin{aligned}
\min_{Q} \|e\|_2^2 &= \min_{Q} \|(1-\hat{P}Q)\hat{v}\|_2^2 \\
\|(1-\hat{P}Q)v\|_2^2 &= \|(1-\hat{P}_A\hat{P}_MQ)\hat{v}_A\hat{v}_M\|_2^2 \\
&= \|(1-z\hat{P}_A\hat{P}_Mz^{-1}Q)\hat{v}_M\|_2^2 \\
&= \|z\hat{P}_A(z^{-1}\hat{P}_A^{-1}-\hat{P}_Mz^{-1}Q)\hat{v}_M\|_2^2 \\
&= \|\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}-\hat{P}_MzQ\hat{v}_M\|_2^2 \\
&= \|\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2}+\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2^\perp}-\hat{P}_Mz^{-1}Q\hat{v}_M\|_2^2 \\
&= \|\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2^\perp}\|_2^2+\|\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2}-\hat{P}_Mz^{-1}Q\hat{v}_M\|_2^2 \\
\min_{\hat{Q}} \|(1-\hat{P}\hat{Q})\hat{v}\|_2^2 &= \min_{\hat{Q}} \|\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2}-\hat{P}_Mz^{-1}Q\hat{v}_M\|_2^2 \Leftrightarrow \\
\hat{Q}_{H_2} &= z(\hat{P}_M\hat{v}_M)^{-1}\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2}
\end{aligned}$$

The minimum error norm is given by

$$\min\|e\|_2^2 = \|\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_{H_2^\perp}\|_2^2.$$

Thus the right half plane zeros pose limits on the minimum achievable error norm. No controller can be designed to overcome these limitations. Toker *et al.* (1997) extend this result to the multivariate case taking into account the effect of input directionality on multivariable performance. They show that even for non-minimum phase systems one can achieve perfect tracking if the generalized step input is aligned in a direction orthogonal to the zero directions. In other words, you can perfectly track only certain types of setpoints and not any setpoint. We state their results for the sake of completeness. A non-minimum phase, right invertible transfer function matrix $P(z)$ can be factorized as

$$P(z) = P_A(z)P_M(z)$$

where $P_A(z)$ is the all pass factor containing all the non-minimum phase zeros of $P(z)$ and $P_M(z)$ has is the minimum phase part. If $a_i, i = 1,..,k$ are the non-minimum phase zeros, then one specific factorization is given by

$$P_A(z) = \prod_{i=1}^{k} P_{A,i}(z), P_{A,i}(z) \triangleq \frac{1-a_i^*}{1-a_i}\frac{z-a_i}{1-a_i^*z}\eta_i\eta_i^H + U_iU_i^H.$$

Here η_i are the unitary vectors obtained by factorizing the zeros one at a time, and U_i are matrices which together with η_i form a unitary matrix; η_i^H denotes the complex conjugate transpose.

Theorem 6.2. *Let u be the step input, $u(k) = v, k \geq 0$, where v is a constant vector with $||v||_2 = 1$. Assume that $P(z)$ is stable, $P(z)$ has full row rank for at least one z and $v \in$ range space of $P(1)$. Then*

$$\min ||e||_2^2 = \sum_{i=1}^{k} \frac{|a_i|^2 - 1}{|a_i - 1|^2} \cos^2 \angle(\eta_i, v)$$

where $\angle(\eta_i, v) = |\eta_i^H v|$.

This results implies that perfect tracking is achievable for MIMO nonminimum phase systems. Consider the case where the plant has a single nonminimum phase zero, a with a zero direction η. The minimum tracking error reduces to

$$\min ||e||_2^2 = \frac{|a|^2 - 1}{|a - 1|^2} \cos^2 \angle(\eta, v).$$

If the step input is aligned with a direction orthogonal to the zero direction, it follows that perfect tracking can be achieved.

6.3.3 Effect of hard constraints

One of the main objectives of a control system, in a practical setting, is to maintain the controlled and the manipulated variables in a feasible region. Physical limitations on the manipulated variables have to be respected and economic considerations require the controlled variables to be maintained at certain desired values. Let us consider the following constraints on the inputs:

$$\underline{u} \leq u(k) \leq \bar{u}.$$

We assume, without loss of generality, that $\underline{u} = -\bar{u}$. In terms of the ∞-norm we can then write

$$||u||_\infty \leq \bar{u}.$$

For the case of deterministic inputs and no model plant mismatch the input in time domain is given by:

$$\begin{aligned} u(k) &= -\sum_i Q(i) r(k-i) \Leftrightarrow \\ ||u||_\infty &\leq ||Q||_1 ||r||_\infty. \end{aligned}$$

Therefore the following condition can guarantee satisfaction of input constraints

$$\begin{aligned} ||Q||_1 ||r||_\infty &\leq \bar{u} \Leftrightarrow \\ ||Q||_1 &\leq \frac{\bar{u}}{||r||_\infty} \Rightarrow ||u||_\infty \leq \bar{u}. \end{aligned} \tag{6.7}$$

Now let us consider constraints on the output $||\hat{y}||_\infty \leq \bar{y}$. For the case where there is no MPM and the only external inputs are setpoint changes:

$$\begin{aligned} \hat{y} &= \hat{T}r \Leftrightarrow \\ \hat{y}(k) &= \sum_i \hat{T}(i)r(k-i) \Leftrightarrow \\ ||\hat{y}||_\infty &\leq ||\hat{T}||_1 ||r||_\infty. \end{aligned}$$

Therefore we can impose the constraints on the complementary sensitivity function,

$$||\hat{T}||_1 \leq \frac{\bar{y}}{||r||_\infty} \Rightarrow ||\hat{y}||_\infty \leq \bar{y}. \tag{6.8}$$

Similarly we can derive bounds for the satisfaction of constraints on the input moves - $\Delta u(k) = u(k) - u(k-1) = (1 - z^{-1})\, \Delta u$,

$$||(1 - z^{-1})Q||_1 \leq \frac{\Delta u_{\max}}{||r||_\infty} \Rightarrow ||\Delta u||_\infty \leq \Delta u_{\max}. \tag{6.9}$$

The above analysis reveals whether a linear controller will satisfy the design constraints on the inputs and the outputs. Alternatively one could design a linear controller which satisfies the above constraints on the 1-norm. Such a controller will guarantee constraint satisfaction although the design procedure will become more complex. The l_1 optimal control approach of Dahleh and Diaz-Bobillo (1995) is based on minimizing the 1-norm of the relevant closed loop transfer function under the above constraints.

The problem of quantifying the effect of constraints on the performance of a model predictive controller is not a straightforward task. In chapter 2 we pointed out that the constraint set for the inputs, u_k is, in general, a time-varying convex set Ω_k. It is well known that the constrained performance can be no better than the unconstrained performance and this gives us a lower bound which can be analytically calculated (see chapter 5). The active set of constraints is impossible to predict and therefore the exact loss in performance due to constraints is an intractable quantity. The worst case loss in performance could be established if for a known set of external inputs we knew which active set of constraints led to highest cost function, *i.e.*,

$$\begin{aligned} J_k^{wc} &= \max_{u_k \in \Omega_k} J_k = \min_{u_k \in \Omega_k} -J_k \Leftrightarrow \\ J_k^{uc} &\leq J_k^* \leq J_k^{wc} \end{aligned}$$

where J_k^{uc} denotes the unconstrained solution at sampling instant k and J_k^* is the optimal constrained solution.

6.4 Effect of uncertainty in the plant model

In practice many components of the model, including the time delay, are never precisely known. For example, the process time delay may be time varying due to changing flow rates. Furthermore most chemical processes are inherently nonlinear, and can be approximated by fixed linear models only in the neighbourhood of the operating point. These and other factors lead to the presence of *uncertainty* in modelling. A controller whose design is based on the nominal model of the process is unlikely to deliver the design performance in the presence of model plant mismatch (MPM). The presence of uncertainty leads to a gap between the designed and the achieved performance. The knowledge of the contribution of uncertainty towards the possible deterioration of controller performance is vital to performance assessment and diagnosis (Kozub, 1996).

The effect of uncertainty on closed loop stability is well known and is the topic of several books and articles (see for example - Doyle, 1992; Morari and Zafiriou, 1989). The objective of this work is to quantify the effect of modelling uncertainty on the achieved performance of a linear model based controller. For the regulatory case with stochastic inputs, the use of Parseval's theorem is instrumental in obtaining bounds on the achieved input and output variances, relative to the design case. For the case of deterministic inputs, the induced H_∞ norm definition is used to derive bounds on the achieved performance in relation to the design performance, with the 2-norm (sum of squared errors (SSE)) as a measure of performance. A simulation example is used to illustrate the effects of uncertainty in time delay on controller performance, in the context of minimum variance benchmark as a measure of performance.

Uncertainty can be structured or unstructured. Two common ways of representing unstructured uncertainty are given below ($\hat{P}$ is the model, P is the plant).

$$\begin{aligned} \text{Additive Uncertainty} &: P = \hat{P} + \Delta \\ \text{Multiplicative Uncertainy} &: P = \hat{P}(1 + \Delta_m) \end{aligned} \tag{6.10}$$

The perturbations belong to the set of stable transfer functions that preserve properness of the P. A practical way of representing the uncertainty is through appropriate weighting functions,

$$\begin{aligned} P &= \hat{P} + W_1\Delta \\ P &= \hat{P}(1 + W_2\Delta_m) \end{aligned} \tag{6.11}$$

where W_1, W_2 are chosen such that $||\Delta||_\infty \leq 1, ||\Delta_m||_\infty \leq 1$. These weighting functions describe the distribution of the model mismatch over the frequency range. In the remaining discussion we consider the case $W_1 = W_2 = 1$ and the subsequent results are derived for a specific perturbation. The results can be easily obtained in terms of the weighting functions for the uncertain set described by 6.11. In this section we analyze the effect of modelling uncertainty on the performance of a control system in terms of the 2-norm. In particular we derive bounds on the achieved performance in terms of the design performance and the ∞-norm of a certain transfer function. The achieved sensitivity and the complementary sensitivity functions, in the presence of the additive uncertainty, are given by

$$S = (I - Q\hat{P})(I + Q\Delta)^{-1}, T = Q\hat{P}(I + Q\Delta)^{-1}. \tag{6.12}$$

Thus the achieved sensitivity function is related to the design sensitivity function by

$$S = \hat{S}(I + Q\Delta)^{-1}. \tag{6.13}$$

In the remaining discussion we restrict ourselves to the case where there are no setpoint changes *i.e.* the regulatory case (see figure 6.2). Using Parseval's Theorem we have for the SISO case:

$$\begin{aligned} y &= SNa \\ Var(y) &= \frac{\sigma_a^2}{2\pi}\int_0^{2\pi} |S|^2|N|^2 dw \\ &= \frac{\sigma_a^2}{2\pi}\int_0^{2\pi} |\hat{S}(1+Q\Delta)^{-1}|^2|N|^2 dw \\ &\leq \frac{\sigma_a^2}{2\pi}||(1+Q\Delta)^{-1}||_\infty^2 \int_0^{2\pi} |\hat{S}|^2|N|^2 dw \\ &\leq ||(1+Q\Delta)^{-1}||_\infty^2 Var(\hat{y}). \end{aligned} \tag{6.14}$$

In a similar way we get the following lower bound,

$$Var(y) \geq \frac{1}{||1+Q\Delta||_\infty^2} Var(\hat{y}).$$

Therefore in time domain the loss in performance due to modeling uncertainty is given by the following performance bounds,

$$\frac{1}{||1+Q\Delta||_\infty^2} \leq \frac{Var(y)}{Var(\hat{y})} \leq ||(1+Q\Delta)^{-1}||_\infty^2. \tag{6.15}$$

The performance bound can be further simplified and expressed in terms of $||Q\Delta||_\infty$ for the case where $||Q\Delta||_\infty < 1$, it can be shown that (see Boyd *et al.*, 1992)

$$\frac{1}{(1+||Q\Delta||_\infty)^2} \leq \frac{Var(y)}{Var(\hat{y})} \leq \frac{1}{(1-||Q\Delta||_\infty)^2}. \tag{6.16}$$

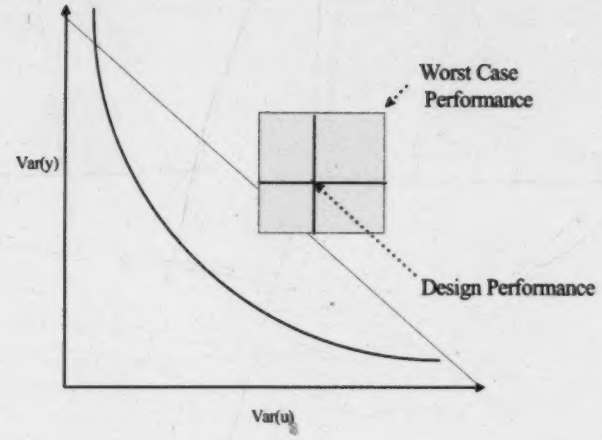


Figure 6.3: *Illustration of the effect of modelling uncertainty on achieved performance*

We can define an upper bound for the increase in input energy (variance) due to uncertainty via a similar derivation.

$$\begin{aligned} u &= -\frac{QN}{1+Q\Delta}a \\ Var(u) &= \frac{\sigma_a^2}{2\pi}\int_0^{2\pi} |Q(1+Q\Delta)^{-1}|^2|N|^2 dw \\ &\leq \frac{\sigma_a^2}{2\pi}||(1+Q\Delta)^{-1}||_\infty^2 \int_0^{2\pi} |Q|^2|N|^2 dw \\ &\leq ||(1+Q\Delta)^{-1}||_\infty^2 Var(\hat{u}). \end{aligned}$$

Once again we have the following time domain bounds on input variance due to MPM

$$\frac{1}{||1+Q\Delta||_\infty^2} \leq \frac{Var(u)}{Var(\hat{u})} \leq ||(1+Q\Delta)^{-1}||_\infty^2. \tag{6.17}$$

Thus we have arrived at a single performance degradation measure for both the input and output variances. The geometry of this region is highlighted in figure 6.3 in relation to the LQG benchmark. If the MPM was negligible the shaded region would be very small and the larger it is the bigger the possible degradation of the control system in terms of loss in performance.

Remark 6.1. *It should be noted here that this is an upper bound and therefore a conservative estimate. The actual loss in performance would depend on the spectrum of the disturbances in relation to the spectrum of the uncertainty and the controller.*

Remark 6.2. *The lower bounds derived in equations 6.16 and 6.17, being less than unity, do not guarantee that the performance is worse in presence of model plant mismatch. This is an interesting observation since it implies that the achieved performance can actually be better than the design performance.*

We can also express similar bounds on the achieved objective function of a linear quadratic guassian (LQG) optimal controller:

$$\begin{aligned} & Var(y) + \lambda Var(u) \\ = \; & \frac{\sigma_a^2}{2\pi}\left\{\int_0^{2\pi} |\hat{S}(1+Q\Delta)^{-1}|^2|N|^2 dw + \int_0^{2\pi} |Q(1+Q\Delta)^{-1}|^2|N|^2 dw\right\} \\ \leq \; & ||(1+Q\Delta)^{-1}||_\infty^2 \left(Var(\hat{y}) + \lambda Var(\hat{u})\right) \;\Leftrightarrow \end{aligned}$$

$$\frac{J_{ach}}{J_{LQG}} \leq ||(1+Q\Delta)^{-1}||_\infty^2. \tag{6.18}$$

In case the above performance bounds indicate a significant MPM, the next step is to determine the robustness of the existing controller with respect to the estimated uncertainty. Let us assume that the controller design was done with the nominal model as the plant. A controller C is said to provide *robust stability* if it provides stability for every plant in the set. The following theorem provides us with a way of assessing robust stability.

Theorem 6.3. *For the multiplicative uncertainty model, the controller C provides robust stability iff $||\hat{T}\Delta_m||_\infty < 1$, where T is the nominal complementary sensitivity function*

Proof. For the proof see Doyle *et al.* (1992).

This is also known as the Small Gain Theorem (SGT). A simple measure of robustness follows from application of the SGT for the additive uncertainty case. In terms of the additive uncertainty, for the system to be robustly stable we have

$$||Q\Delta||_\infty < 1. \tag{6.19}$$

Let

$$\rho \triangleq ||Q\Delta||_\infty : \text{ Robustness Measure} \tag{6.20}$$

The smaller the value of ρ the more robust is the control system, and conversely, the closer it is to unity the closer it is to instability due to MPM. In terms of the robustness measure the performance measure can now be rewritten as:

$$\frac{1}{(1+\rho)^2} \leq \frac{Var(y)}{Var(\hat{y})} \leq \frac{1}{(1-\rho)^2}. \tag{6.21}$$

Therefore the smaller the value of $\rho(\ll 1)$,the achieved performance will be close to the design performance. It is well known that the design performance itself is lower for a system with a large robustness margin. On the other hand if the feedback system has poor robustness properties, *i.e.* $\rho \approx 1$, the higher will be the uncertainty in performance *i.e.* we pay the price for being less robust by a possible loss in the design performance. There is a higher risk of instability if there is a further change in the plant dynamics. The above discussion portrays the traditional trade-off between performance and robustness in a different light. Note that $y, \hat{y}$ are themselves functions of $Q, \Delta, \hat{P}$. In general, an estimate of the uncertainty may not be available. Appendix A gives a way of estimating the multiplicative uncertainty under closed loop conditions via non-parametric identification techniques.

6.4.1 Deterministic Inputs

In a practical setting a process is subject to different kinds of inputs such as setpoint changes, random walk type disturbances, measurement noise, *etc.* In this section we consider the effect of modeling uncertainty in the presence of deterministic as well as stochastic inputs. Let us take the case where we have setpoint changes, *i.e.*, the tracking performance,

$$\begin{aligned} y &= Tr \\ e &= r - y = (1-T)r = Sr \\ e &= \hat{S}(1+Q\Delta)^{-1}r \Leftrightarrow \\ \|e\|_2^2 &\leq \|(1+Q\Delta)^{-1}\|_\infty^2 \|\hat{e}\|_2^2. \end{aligned}$$

Thus the we have a worst case bound on the sum of the squared errors (2-norm) using the same transfer function $(1+Q\Delta)^{-1}$ as before. The same limit can be re-derived for the case where we have deterministic disturbances

$$\begin{aligned} e &= S(r+d) \Leftrightarrow \\ \|e\|_2^2 &\leq \|(1+Q\Delta)^{-1}\|_\infty^2 \|\hat{e}\|_2^2. \end{aligned}$$

6.4.2 Mixed Inputs

Now let us consider the case where we have both, deterministic as well as stochastic disturbances,

$$\begin{aligned} y &= Tr + Sd + SNa \\ e &= Sr + Sd + SNa \\ &= (1 + Q\Delta)^{-1}\hat{S}(r + v + Na) \Leftrightarrow \\ \|e\|_2^2 &\leq \|(1 + Q\Delta)^{-1}\|_\infty^2 \|\hat{e}\|_2^2. \end{aligned} \tag{6.22}$$

6.4.3 Uncertainty in the time delay

Here we discuss the special case where the uncertainty is present in the process time delay. In practice uncertainty in the process dead time is very common and it is a useful exercise to determine the effect of this uncertainty on controller performance. First we re-derive the results in the previous section in terms of the multiplicative uncertainty model. The multiplicative uncertainty is particularly well suited for representing variable delay. The multiplicative uncertainty is of the form

$$P = \hat{P}(1 + \Delta_m).$$

The multiplicative uncertainty is related to the additive uncertainty by

$$\Delta_m = \hat{P}^{-1}\Delta. \tag{6.23}$$

The design complementary sensitivity function is given by

$$\hat{T} = Q\hat{P}.$$

Therefore we have,

$$\begin{aligned} 1 + Q\Delta &= 1 + Q\hat{P}\hat{P}^{-1}\Delta \\ &= 1 + \hat{T}\Delta_m. \end{aligned}$$

The uncertainty in time delay can be characterized via the multiplicative uncertainty in the following way,

$$\begin{aligned} P &= \hat{P}z^{-\theta}, \theta_{\min} \leq \theta \leq \theta_{\max} \\ &= \hat{P}(1 + z^{-\theta} - 1) \end{aligned} \tag{6.24}$$

hence we have,

$$\begin{aligned} \Delta_m &= z^{-\theta} - 1 \\ \Delta_m(e^{jw}) &= e^{-j\theta w} - 1 \\ &= \cos(\theta w) - j\sin(\theta w) - 1 \\ |\Delta_m(e^{jw})| &= |\cos(\theta w) - j\sin(\theta w) - 1| \\ &= \sqrt{2 - 2\cos(\theta w)}. \end{aligned} \tag{6.25}$$

For the worst case uncertainty in the delay, $\theta = \theta_{\max}$, $|\Delta_m(e^{jw})|$ is a periodic function of frequency and takes the maximum values at the frequencies, $w = \frac{n\pi}{\theta_{\max}}, |\Delta_m(e^{jw})| = 2$. Using this property the bounds on the worst case performance can be derived as:

$$\begin{aligned} \frac{1}{\left(1 + ||\hat{T}\Delta_m||_\infty\right)^2} &\leq \frac{Var(y)}{Var(\hat{y})} \leq \frac{1}{\left(1 - ||\hat{T}\Delta_m||_\infty\right)^2} \Leftrightarrow \\ \frac{1}{\left(1 + ||\hat{T}||_\infty||\Delta_m||_\infty\right)^2} &\leq \frac{Var(y)}{Var(\hat{y})} \leq \frac{1}{\left(1 - ||\hat{T}||_\infty||\Delta_m||_\infty\right)^2} \Leftrightarrow \\ \frac{1}{\left(1 + 2||\hat{T}||_\infty\right)^2} &\leq \frac{Var(y)}{Var(\hat{y})} \leq \frac{1}{\left(1 - 2||\hat{T}||_\infty\right)^2} \end{aligned} \tag{6.26}$$

if $||\hat{T}||_\infty < 1/2$. This condition may not be satisfied in practice and as a result the simplified expression above will be applicable in limited situations. Notice that the above expression has only the complementary sensitivity function appearing in the upper and lower bounds. The uncertainty expression does not appear explicitly in the above inequality, it is true irrespective of what the actual magnitude of the uncertainty in the delay is. It should be noted here that these are very conservative bounds. Similar conditions can also be derived for the case of gain mismatch and the case of joint mismatch in the delay and gain. These derivations are omitted here for the sake of brevity.

6.4.4 Extension to the multivariate case

For the multivariable case similar results can be derived based on the definition of the 2-norm.

$$
\begin{aligned}
\hat{S}(I+Q\Delta)^{-1} &= (I+\Delta Q)^{-1}\hat{S} \qquad (6.27)\\
y &= (I+\Delta Q)^{-1}\hat{S}Na\\
y &= (I+\Delta Q)^{-1}\hat{y}\\
\frac{1}{\|I+\Delta Q\|_\infty^2} &\le \frac{\|y\|_2^2}{\|\hat{y}\|_2^2} \le \|(I+\Delta Q)^{-1}\|_\infty^2\\
\frac{1}{(1+\bar{\sigma}(\Delta Q))^2} &\le \frac{\|y\|_2^2}{\|\hat{y}\|_2^2} \le \frac{1}{(1-\bar{\sigma}(\Delta Q))^2},\\
\text{similarly } \frac{1}{(1+\bar{\sigma}(\Delta Q))^2} &\le \frac{\|u\|_2^2}{\|\hat{u}\|_2^2} \le \frac{1}{(1-\bar{\sigma}(\Delta Q))^2}
\end{aligned}
$$

where

$$\bar{\sigma}(G(e^{jw})) = \sup_w \sigma_i(G(e^{jw})).$$

$\sigma_i(G(e^{jw}))$ are the singular values of G defined via the singular value decomposition (Morari and Zafiriou, 1989).

Uncertainty in the Interactor: The Diagonal Case

Here we consider the case of the diagonal interactor for the 2x2 case. The plant is given by

$$P = \hat{P}\begin{bmatrix} z^{-d_1} & \\ & z^{-d_2} \end{bmatrix}$$

where d_1, d_2 are is the maximum change in the original time delays for the two input channels. Therefore we have,

$$
\begin{aligned}
P &= \hat{P}(I+D-I) = \hat{P}(I+\Delta_m)\\
\Delta_m &= \begin{bmatrix} z^{-d_1}-1 & \\ & z^{-d_2}-1 \end{bmatrix}.
\end{aligned}
$$

The frequency domain bound on the uncertainty in the interactor is given by

$$
\begin{aligned}
\Delta_m(e^{-jw}) &= \begin{bmatrix} e^{-jwd_1}-1 & \\ & e^{-jwd_2}-1 \end{bmatrix} \qquad (6.28)\\
&= I\begin{bmatrix} |\Delta_1| & \\ & |\Delta_2| \end{bmatrix}\begin{bmatrix} \angle\Delta_1 & \\ & \angle\Delta_2 \end{bmatrix}\\
\sigma(\Delta_m(e^{-jw})) &= \begin{bmatrix} |\Delta_1| & \\ & |\Delta_2| \end{bmatrix}.
\end{aligned}
$$

This approach can be extended to higher dimensions assuming that the interactor remains diagonal. In the last section we present a 2×2 simulation example showing the effect of the uncertainty in the interactor on the LQG optimal cost function.

6.4.5 Uncertainty in the noise model

A similar analysis is possible for a mismatch in the noise model. Uncertainty in the noise model affects closed loop performance. Let Δ_m^N be the multiplicative uncertainty in the noise model,

$$N = \hat{N}(1 + \Delta_m^N) \tag{6.29}$$

form of uncertainty in the noise model can be used to represent a wide range of disturbances through a single *uncertain* noise model. Stochastic realization theory (Astrom and Wittenmark, 1991) allows one to realize different kinds of disturbance signals as filtered white noise processes. Thus a class of inputs could be represented by uncertainty bounds on a fixed nominal noise model $\hat{N}$. For the case where there is no uncertainty in the plant model, the maximum contribution of Δ_m^N to increasing output variance is given by

$$\begin{aligned} Var(y) &= \frac{\sigma_a^2}{2\pi}\int_{-\infty}^{\infty} |S|^2|N|^2 dw \qquad (6.30)\\ &= \frac{\sigma_a^2}{2\pi}\int_{-\infty}^{\infty} |S|^2|\hat{N}(1+\Delta_m^N)|^2 dw \\ &\leq \frac{\sigma_a^2}{2\pi}||(1+\Delta_m^N)||_\infty^2 \int_{-\infty}^{\infty} |\hat{S}|^2|\hat{N}|^2 dw \\ &\leq ||(1+\Delta_m^N)||_\infty^2 Var(\hat{y}). \end{aligned}$$

Thus the combined effect of uncertainty in the noise and plant models can be expressed as

$$\frac{Var(y)}{Var(\hat{y})} \leq ||(1+\Delta_m^N)||_\infty^2 ||(1+Q\Delta)^{-1}||_\infty^2. \tag{6.31}$$

6.5 Impact of nonlinearity on performance

Chemical processes are inherently nonlinear. Several authors have proposed measures for quantifying the nonlinearity of a process - Guay *et al.* (1995), Allgower (1995), Nikolaou (1993) and Stack and Doyle (1997). Guay *et al.* (1995) used the curvature of the nonlinear mapping to obtain a nonlinearity index over a specified operating region. Nikolaou (1993) defines a nonlinearity measure based on the development of a nonlinear operator 2-norm via a carefully constructed inner product for nonlinear operators. The residuals from the best linear approximation are used to quantify the nonlinearity over a particular input interval. Allgower (1995) used the following operator based measure to quantify nonlinearity,

Definition 6.1. *The nonlinearity measure ϕ_F^{Ω} of a BIBO-stable, causal I/O system N : $R^{n_u} \mapsto R^{n_y}$ for input signals $u \in \Omega \subseteq R^{n_u}$ is defined by the nonnegative number*

$$\phi_F^{\Omega} = \inf_{P \in \mathcal{P}} \sup_{u \in \Omega} \frac{||F(u) - P(u)||}{||u||} \tag{6.32}$$

with $P : R^{n_u} \mapsto R^{n_y}$ being a linear operator , $||.||$ defines a norm over the relevant input and output spaces.

Allgower (1995) describes a way of parametrizing the input space using sinuosidal signals with different frequencies and obtaining the nonlinearity measure in equation 6.32. The nonlinearity measure is thus defined as the largest deviation between the output of a nonlinear system F and the best linear approximation $\hat{P}$ over a class of inputs, $u \in \Omega$ with respect to a particular norm on the input and output spaces.

In practice a linear controller operating on a nonlinear process is the most common scenario. Here we use the measure defined by Allgower (1995) to obtain an upper bound on the deviation in the design performance due to the intrinsic nonlinearity of the process. Assume that a linear model $\hat{P}$ has been chosen for purposes of designing a linear feedback controller, C. The setpoints, $r \in A_r \subseteq R^{n_y}$ drive the closed loop process. The closed loop relations for the design and the achieved cases are given by

$$\begin{aligned} \textit{Design} \quad &: \quad \hat{y} = \hat{T}r, e = r - \hat{y} = \hat{S}r \\ \textit{Achieved} \quad &: \quad y = F_c(r) = F[\sum_{i=0}^{N} C_i e(k-i)] = F\{\sum_{i=0}^{N} C_i(r(k-i) - y(k-i)\} \end{aligned} \tag{6.33}$$

where F_c denotes the closed loop nonlinear operator relating r to y and C_i are the impulse response coefficients of the linear controller. Assuming that the linear controller stabilizes the nonlinear system we define the following measure of closed loop nonlinearity with respect to the linear closed loop system $\hat{T}$,

$$\phi_{F_c}^{A_r} = \sup_{r \in A_r} \frac{||F_c(r) - \hat{T}(r)||}{||r||} \tag{6.34}$$

assuming that such an upper bound exists. This is a measure of the closed loop nonlinearity of the system which can be lesser or greater than the open loop nonlinearity depending upon the region of inputs in which the controller excites the nonlinear plant. It is also determined by the class of setpoints which drives the feedback system. An application of the triangle inequality on the 2-norm of the achieved tracking errors gives us the following relation,

$$\begin{aligned} ||r - y|| &= ||r - y + \hat{y} - \hat{y}|| \\ &\leq ||r - \hat{y}|| + ||y - \hat{y}|| \\ &\leq ||r - \hat{y}|| + \phi_{F_c}^{A_r} ||r||. \end{aligned} \tag{6.35}$$

Thus we have

$$||e|| \leq ||\hat{e}|| + \phi_{F_c}^{A_r}||r||$$

i.e. the maximum deviation of the achieved closed performance from the linear closed loop performance is function of the closed loop nonlinearity scaled by the by the 2-norm of the reference signal.

6.5.1 Effect of mismatch on model predictive controllers

The development of the design case measure for model predictive controller in Chapter 5 proves instrumental in quantifying the effect of nonlinearity on MPC performance. Recalling the inequality for the achieved objective function

$$|\hat{J}_k - ||\hat{y}_k - y_k||_\Gamma^2 - ||\Delta u_k - \Delta\hat{u}_k||_\Lambda^2| \leq J_k \leq \hat{J}_k + ||\hat{y}_k - y_k||_\Gamma^2 + ||\Delta u_k - \Delta\hat{u}_k||_\Lambda^2. \quad (6.36)$$

The achieved performance is significantly affected by the control relevant mismatch term

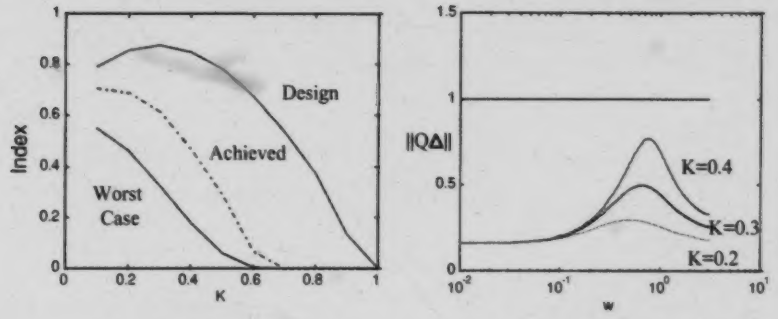
$$||\hat{y}_k - y_k||_\Gamma^2$$

which consists of the modelling errors weighted and squared over the prediction horizon under closed loop conditions. This mismatch is a function of the inherent nonlinearity of the process, the model predictive controller and the external inputs, under feedback. It is important to note that under feedback $||\Delta u_k - \Delta\hat{u}_k||_\Lambda^2$ will be strongly correlated to $||\hat{y}_k - y_k||_\Gamma^2$. The effect of this control relevant mismatch is often suppressed by ensuring $\hat{J}_k >> ||\hat{y}_k - y_k||_\Gamma^2 + ||\Delta u_k - \Delta\hat{u}_k||_\Lambda^2$. This is equivalent to relaxing the design requirements and can be achieved by increasing the move suppression factors Λ in the design objective functions.

6.6 Simulation examples

Example 1. Here we revisit the simulation example from Desborough and Harris (1992) (see Chapter 5). The nominal transfer functions are given by

$$\begin{aligned} Q &= C(1+\hat{P}C)^{-1} = \frac{K}{(1-z^{-1})}\frac{(1-z^{-1})}{(1-z^{-1}+Kz^{-2})} \\ &= \frac{K}{(1-z^{-1}+Kz^{-2})} \\ \hat{T} &= Q\hat{P} = \frac{Kz^{-2}}{(1-z^{-1}+Kz^{-2})}. \end{aligned}$$



Performance (a) Robustness (b)

Figure 6.4: *The performance and robustness properties of the control system with a unit delay mismatch.*

An additional delay was introduced in the *plant*,

$$y(k) = u(k-3) + \frac{1-0.2z^{-1}}{1-z^{-1}}a(k).$$

Thus we have the multiplicative uncertainty of the form

$$\Delta_m = z^{-1} - 1.$$

Figure 6.4 shows the effect of the time delay mismatch on performance and stability of the designed controller. Figure 6.4 (a) shows the achieved performance index in relation to the nominal performance for different controller gains. The performance index is based on the minimum variance benchmarking (Harris, 1989). The minimum achievable variance for the process with two sample delays is given by

$$\sigma_{mv}^2 = (1+0.8^2)\sigma_a^2 = 1.64\sigma_a^2.$$

The performance index is then calculated as a ratio of the minimum achievable variance to the achieved variance,

$$\mu = \frac{\sigma_{mv}^2}{\sigma_y^2}.$$

The effect of uncertainty on the MVC performance index is quantified as follows:

$$\begin{aligned} \frac{\sigma_{mv}^2}{\sigma_y^2} &= \frac{\sigma_{mv}^2}{\sigma_{\hat{y}}^2}\frac{\sigma_{\hat{y}}^2}{\sigma_y^2} \Leftrightarrow \\ \frac{\hat{\mu}}{\|(1+\hat{T}\Delta_m)^{-1}\|_\infty^2} &\leq \mu \leq \hat{\mu}(\|1+\hat{T}\Delta_m\|_\infty^2). \end{aligned} \tag{6.37}$$

The design curve in figure 6.4 (a) was based on the performance index for the nominal case and the achieved curve corresponds to the unit sample delay mismatch. The worst case bound was derived using equation 6.15 . The controller shows a significant degradation in performance due to an increase in the time delay of a sample interval. Figure 6.4 (b) shows the robustness properties of the integral controller for a unit sample time delay mismatch. As the controller gain increases, the robustness margin deteriorates. For K=0.2, the controller is highly robust and the achieved performance is close to the design performance (see figure 6.4 (a)). For K=0.4 the robustness margin is small and the achieved performance is significantly different from the design performance. Figure 6.5 (a) shows the achieved and the design servo response when the time delay is underestimated. The achieved performance is worse than the design performance,

$$\frac{||e||_2^2}{||\hat{e}||_2^2} = 1.8117 \leq ||(1+\hat{T}\Delta_m)^{-1}||_\infty^2 = 4.84.$$

We also consider the case where the delay is overestimated *i.e.* the delay is assumed to be 2 units when the actual delay is 1 unit. Figure 6.5 (b) shows the achieved and design closed loop step responses for the case of time delay overestimation ($\Delta_m = z^1 - 1$),

$$\frac{||e||_2^2}{||\hat{e}||_2^2} = 0.6428 \geq \frac{1}{||(1+\hat{T}\Delta_m)||_\infty^2} = 0.3486.$$

The perturbation considered here, despite being non-causal, is valid since it preserves the properness of the closed loop transfer functions.

It is interesting to note that when the time delay is overestimated the performance improves over the design case whereas when the delay is underestimated the performance deteriorates. Thus this example serves to illustrate the point that achieved performance can improve even in the presence of mismatch and both the upper and lower limits in equation 6.15 are meaningful.

Constraints: Let us consider the following constraints on the inputs: $|u(k)| \leq 1$. The 1-norms of the controller transfer function for increasing values of K are shown in figure 6.6. It follows from equation 6.7 that the input constraints will be satisfied for step change in the setpoint ($||r||_\infty = 1$) up to a controller gain $K = 0.2$. Beyond that there is no guarantee that the input constraints will be satisfied by the controller. Figure 6.7 shows the input profiles for a step change in setpoint. For $K = 0.3, 0.4$ notice the constraint violations. For $K = 0.2$ the input stays in the feasible region as predicted by equation 6.7.

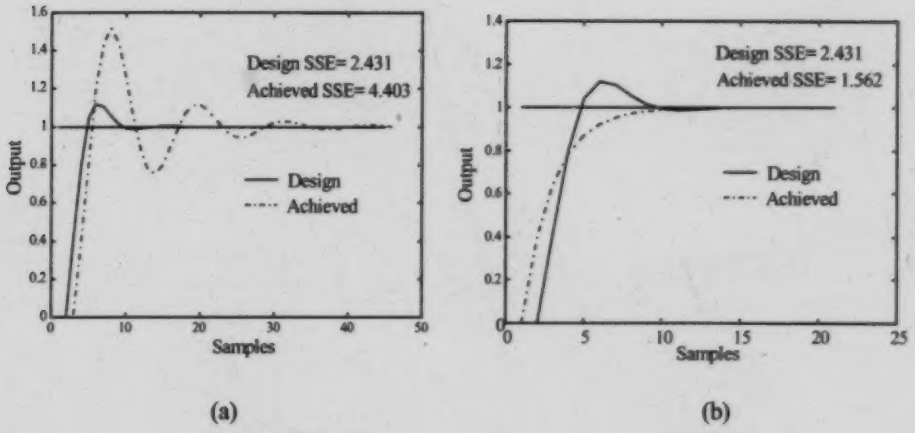


Figure 6.5: *Comparison of achieved and designed closed loop step responses for the case where (a) the delay is underestimated and (b) the delay is overestimated.*

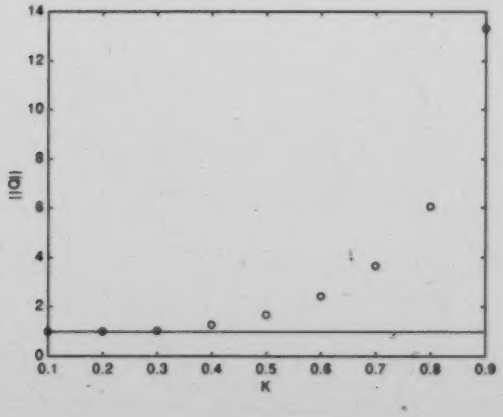


Figure 6.6: *The 1-norm of the controller transfer function Q, against increasing controller gains.*

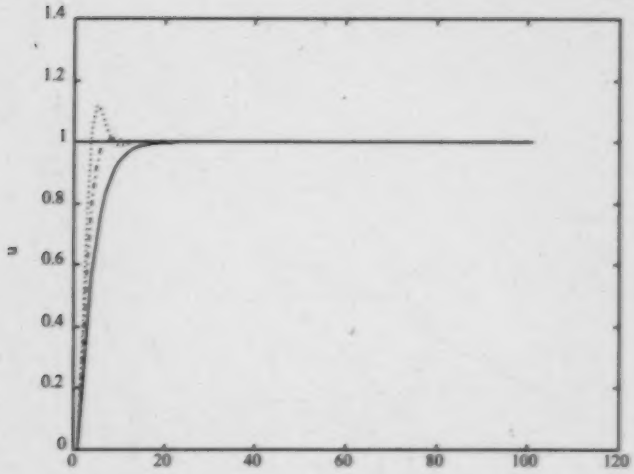


Figure 6.7: *The controller response (u) for K=0.2 (solid line), K=0.3 (dash-dotted line) and K=0.4 (dotted line)*

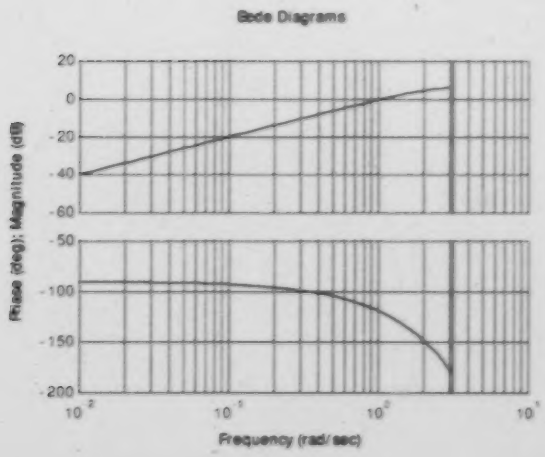


Figure 6.8: *The phase and magnitude characteristics of the (multiplicative) uncertainty in time delay.*

Example 2. Let us consider the example from Bannerjee (1996) again:

$$P = \frac{0.007z^{-1} + 0.0212z^{-2} + 0.0036z^{-3}}{1 - 1.9031z^{-1} + 1.1514z^{-2} - 0.2158z^{-3}}$$

which is approximated by the first order model:

$$\begin{aligned}
\hat{P} &= \frac{0.0419z^{-2} + 0.0719z^{-3}}{1 - 0.8969z^{-1}} \\
\hat{P}_A\hat{P}_M &= \frac{z^{-2}(0.5828 + z^{-1})}{(1 + 0.5828z^{-1})} \frac{0.0719(1 + 0.58228z^{-1})}{(1 - 0.8969z^{-1})}.
\end{aligned}$$

Non-minimum phase zeros: Consider a step change in the setpoint. The following calculations based on Theorem 1, give the minimum error norm for the H_2 optimal controller. Figure 6.9 verifies the tracking properties of the H_2 optimal controller.

$$\begin{aligned}
\hat{v} &= \frac{1}{1 - z^{-1}} = \hat{v}_M \\
\hat{P}_A^{-1}\hat{v}_M &= \frac{(1 + 0.5828z^{-1})}{z^{-2}(0.5828 + z^{-1})} \frac{1}{(1 - z^{-1})} \\
\{z^{-1}\hat{P}_A^{-1}\hat{v}_M\}_* &= \frac{z^{-1}}{(1 - z^{-1})} \\
Q_{H_2} &= \hat{P}_M^{-1} = \frac{(1 - 0.8969z^{-1})}{0.0719(1 + 0.58228z^{-1})} \\
(1 - \hat{P}\hat{Q}_{H_2})\hat{v} &= (1 - \hat{P}_A)\hat{v} \\
&= \left\{1 - \frac{z^{-2}(0.5828 + z^{-1})}{(1 + 0.5828z^{-1})}\right\} \frac{1}{(1 - z^{-1})} \\
&= \frac{1 + 1.5828z^{-1} + z^{-2}}{1 + 0.5828z^{-1}} \\
||(1 - \hat{P}\hat{Q}_{H_2})\hat{v}||_2^2 &= 2.2636
\end{aligned}$$

Constraints: The design complementary sensitivity function has the 1-norm $||\hat{T}||_1 = 2.1656$. Thus for a step change in the setpoint, $||y||_\infty \leq 2.1656$. From figure 6.9, $||y||_\infty = 1.2431$. **Uncertainty:** An unconstrained model predictive controller (MPC) was designed using the first order model with two different control weightings - $\lambda = 0(Q_1)$, $\lambda = 0.5(Q_2)$. For the unconstrained case the predictive control law can be expressed in the form of a transfer function (see chapter 2). Figure 6.10 (b) shows the controller inverses, in the IMC form, as a function of frequency in relation to the additive uncertainty (Δ). Q_2 is seen to be more robust in comparison to Q_1. The achieved variances (normalized) for Q_2 are shown in figure6.10 (a). The following bounds were derived for the achieved variances:

$$\begin{aligned}
||Q\Delta||_\infty &= 0.4386 \Leftrightarrow \\
0.4832 \leq &\frac{Var(y)}{Var(\hat{y})} \leq 3.17
\end{aligned}$$

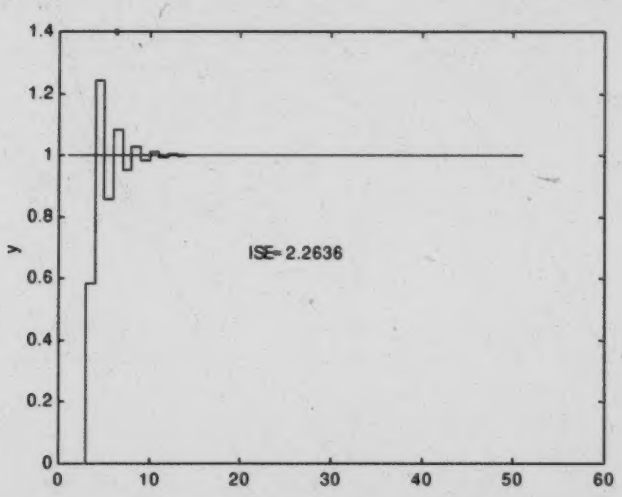


Figure 6.9: *Nominal closed loop step response of a H_2 optimal controller*

Example 3. The Wood-Berry column revisited. Consider the discretized model at the sampling rate of 1 minute,

$$G(z^{-1}) = \hat{D}^{-1}G_0 = \begin{bmatrix} z^{-2} & \\ & z^{-4} \end{bmatrix} \begin{bmatrix} \frac{0.744}{1-0.9419z^{-1}} & \frac{-0.8789}{1-0.9535z^{-1}} \\ \frac{0.5786z^{-4}}{1-0.9123z^{-1}} & \frac{-1.3015}{1-0.9239z^{-1}} \end{bmatrix}.$$

We consider the uncertainty of 1 sample delay in each channel,

$$\sigma(\Delta_m) = \sqrt{2(1-\cos w)}I.$$

A LQG controller was designed with $\Gamma = I, \Lambda = \lambda I$ based on G_0. The optimal cost function for the LQG controller is shown in figure 6.11 along with the uncertainty intervals according to equation 6.18.

The solid line in the center is the nominal cost function J_{LQG} vs. λ. The achieved performance for an actual change in delay is shown by the dashdotted lines and the upper and lower solid lines are the bounds on the achieved cost function. The performance bounds were derived using equation 6.18. Figure 6.12 shows the effect of λ on the robustness of the LQG controller.

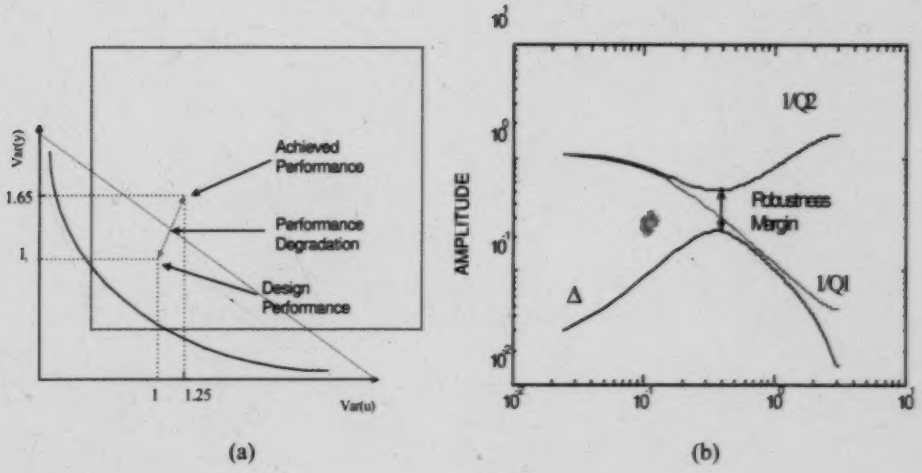


Figure 6.10: *The effect of a structural mismatch on the performance and stability of a MPC controller. (a) The design and achieved performance for Q_1 and (b) the robustness properties of Q_1 and Q_2.*

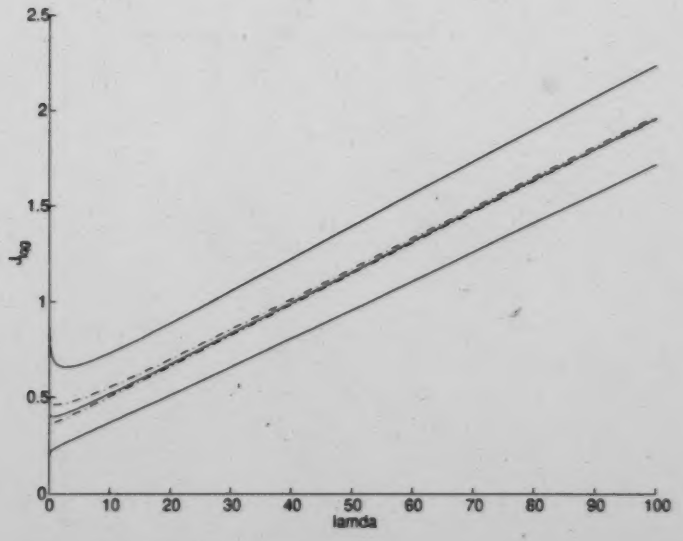


Figure 6.11: *Effect of uncertainty in the interactor on the optimal cost function for LQG.*

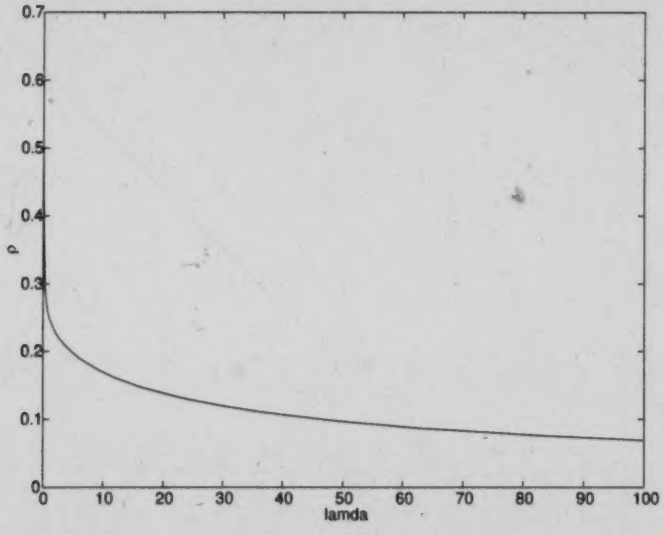


Figure 6.12: *Effect of control weighting on robustness margin.*

6.7 Conclusions

In this work we have considered the contributions of the following factors on the *achievable* and the *achieved* performance of a control system:

1. **Non-Minimum Phase Systems.** Limitations on achievable performance due to nonminimum phase zeros are reviewed for the univariate and multivariate cases. The importance of directionality is emphasized in the result due to Toker *et al.* (1997).

2. **Constraints.** The presence of input and output constraints imposes limits on the 1-norm of the controller and the complementary sensitivity transfer function. This was shown using the induced norm definition of the 1-norm. The exact effect of constraints on the MPC performance is intractable. However, the unconstrained performance provides a useful lower bound.

3. **Modeling Uncertainty.** In this work we have considered the contributions of the modelling uncertainty on the *achieved* performance of a control system. The presence of uncertainty in the modelling step is reflected in uncertain controller performance. Time domain bounds on the achieved performance are obtained based on the knowledge of the uncertainty. As a special case, the effect of varying time delay is captured using multiplicative uncertainty. Similar performance bounds are derived for the case

of deterministic inputs. A simulation example is presented to illustrate the effect of uncertainty in time delay on achieved performance. Similar results are derived for multivariable systems.

4. **Input Uncertainty.** A process is often subject to different kinds of inputs such as step changes in the setpoint, random changes in the unmeasured disturbances, *etc.* The effect of uncertainty in the presence of a combination of deterministic and stochastic inputs is illustrated.

5. **Impact of Nonlinearity.** A closed loop nonlinearity measure as per the definition of Allgower (1995) was shown to bound the achieved performance of a linear model-based system on a nonlinear plant. The choice of a suitable performance cost function in relation to the impact of nonlinearity on model based predictive controllers was highlighted.

Thus the effect of several factors on deteriorating performance has been quantified. Some of these factors such as nonminimum phase zeros pose fundamental limitations *i.e.* they are controller independent while others such as the effect of uncertainty or nonlinearity under closed loop conditions are controller dependent. The task of isolating the contribution of each of these factors to the overall degradation of performance is a nontrivial one. In the next chapter we take a case study approach to performance diagnosis of an industrial model predictive controller to illustrate the complexity of the diagnosis step.

Chapter 7

Performance Analysis of Model Predictive Controllers: Industrial Applications

7.1 Introduction

The first industrial applications of model predictive control were reported by Richalet *et al.* (1978) and Cutler and Ramaker (1980). Richalet *et al.* (1978) called their implementation model predictive heuristic control (MPHC) and the commercial version was known as the identification and command algorithm (IDCOM). Cutler and Ramaker called their version dynamic matrix control (DMC) after the S matrix which they named as the dynamic matrix. Qin and Badgwell (1996) give an excellent survey of the MPC technologies commercially available and compare their merits and demerits. They report the total number of MPC applications to be over 2200, with majority (67%) of the applications being in the refinery area. The DMC corporation has the largest number of applications- over 600- in the area of refining and petrochemicals. The largest single application reported by DMC Co. is an olefins plant with 603 inputs and 283 outputs.

Dynamic matrix control (DMC) which was originally proposed by Cutler and Ramaker in 1980 is a type of model predictive controller (MPC) . In practice, DMC is implemented in two stages:

1. The dynamic error minimization step, which computes the manipulated variables. The design objective function comprises of weighted deviations of the tracking error, weighted control effort and a penalty on the deviation of the final input from the

[1] A version of this chapter was presented as "Performance Analysis of Model-based Predictive Controllers: An Industrial Case Study", 1998, Rohit S. Patwardhan, Sirish L. Shah, Genichi Emoto, and Hiroyuki Fujii, AICHE Annual Meeting, Miami.

steady state targets (Qin and Badgewell, 1996) . The last term is tantamount to having terminal state weighting which is known to guarantee nominal stability (Lee, 1996). The setpoint and the steady state targets are computed by

$$J_{des} = \sum_{i=1}^{p} ||y_{ss} - \hat{y}(k+i/k)||^2_{\Gamma_i} + \sum_{i=1}^{m} ||\Delta u(k+i-1)||^2_{\Lambda_i} + ||u(k+m-1) - u_{ss}||^2_{R_i} \tag{7.1}$$

where the terms are as explained in Chapter 2.

2. The linear programming (LP) stage which deals with steady state optimization and constraint management. A linear program of the following form is solved at every sampling instant (Emoto *et. al*, 1996). The objective function consists of a profit term and a second term which penalized constraint violations on the controlled variables. The constraints are made up of soft constraints on the controlled variables and hard constraints on the manipulated variables.

$$\min \sum_{i=1}^{n_u} c_i \Delta u_{ss,i} + M(\sum_{i=1}^{n_y} w_{\max,i}\lambda_{\max,i} + \sum_{i=1}^{n_y} w_{\min,i}\lambda_{\min,i}) \tag{7.2}$$

subject to

$$\begin{aligned} y_{\max,i} &\geq y_i + \sum_{j=1}^{n_u} K_{ij}\Delta u_{ss,j} - \lambda_{\max,j} \\ y_{\min,i} &\leq y_i + \sum_{j=1}^{n_u} K_{ij}\Delta u_{ss,j} - \lambda_{\min,j}, i = 1, ..., n_y \\ u_{\min,i} &\leq u_i + \Delta u_{ss,i} \leq u_{\max,i}, i = 1, ..., n_u \end{aligned}$$

where $K = [K_{ij}]$: Steady state gain matrix, M: Large weight, λ : Slack variable, w : Penalty weight for constraint violations. A schematic of this control structure is shown in figure 7.1.

Performance analysis of DMC presents a challenging research problem of significant industrial relevance. A constrained DMC type controller is essentially a linear time varying controller, especially when operating at the constraints. The large dimensionality of the problem is also an important issue. Any performance measure for DMC should effectively deal with the multivariate nature of the system and the constraint issue. The main characteristic of a desirable performance measure should be to relate the information contained in a large dimensional data set to a single scalar measure of performance in a transparent manner. Diagnosis of poor performance is of great practical importance. If a performance index reveals poor performance then it is important tc establish the underlying causes and suggest corrective actions to improve performance.

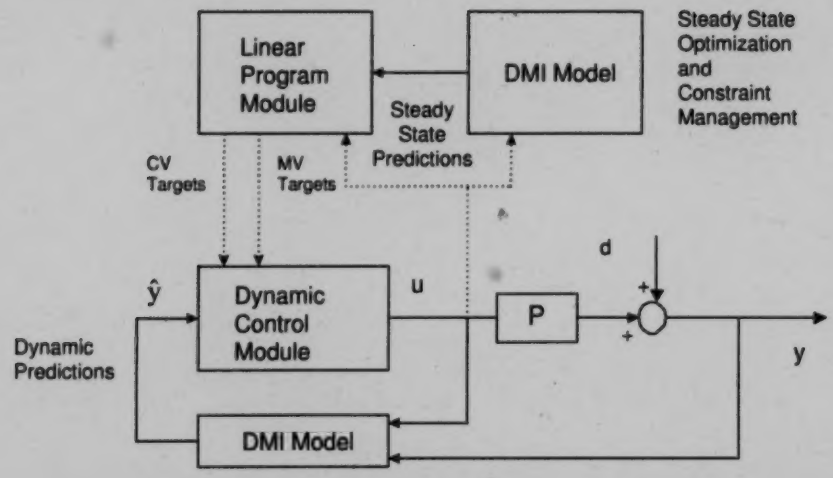


Figure 7.1: *Schematic of the DMC control structure. The controller in the conventional sense is shown as the dynamic control module with the DMI model in the feedback path. The LP part at the top sets the targets for the controlled and the manipulated variables.*

The purpose and contribution of this chapter is to describe an industrial case study involving performance analysis of two DMC applications at the Mizushima plant of the Mitsubishi Chemical Company[2]. A propylene splitter DMC and a de-methanizer DMC were the applications evaluated during this case study. Practical performance assessment and diagnostic techniques were proposed and implemented on the aforementioned applications. The main outcomes were:

- Practical measures to assess DMC performance.
- The role of constraints in analyzing MPC performance.
- Use of closed loop identification to assess the impact of mismatch on MPC type controllers.
- The role of the LP step of DMC in determining closed loop performance and stability.
- An integrated approach to performance assessment which involves checking the lower level PID controllers and upper level RTO.

The broad scope of this work is divided into three tasks:

[2]This project was carried out as part of a 3 month industrial internship program. The author was with the Development and Engineering Research Center at the Mizushima plant, Japan during the course of this work.

1. *How healthy is your control system* ? This is the performance assessment stage. This step is mainly concerned with quantification of performance. One or more performance measures can be used to decide if the control system is delivering satisfactory performance. The first objective of this project was to choose appropriate performance measures for DMC.

2. *Why is your control system in poor health*? This is the diagnostics stage. If the assessment step indicates poor performance, the logical step is to establish the reasons for poor performance. This is a relatively new area of research. To the best of the authors knowledge there is only one publication in open literature addressing this problem in its own right - Kesavan and Lee (1997). Thus the second goal of this project was to devise ways of diagnosing poor performance.

3. *How do we improve performance*? After isolating the causes of poor performance, corrective actions should be taken to restore the health of the control system. The success of this step is heavily dependent on the diagnosis step.

A schematic of the possible choices arising at each step is given in figure 7.2 . The list is by no means exhaustive, it is representative. Several authors have realized the importance of the diagnosis step -Kozub (1993, 1997), Kesavan and Lee (1997).

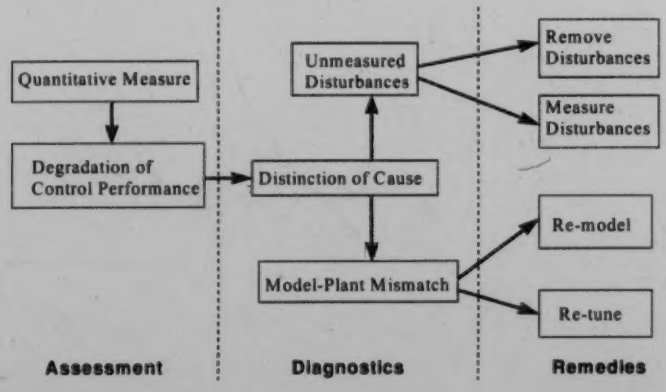


Figure 7.2: *A representative diagram showing the various steps in performance analysis.*

This chapter is organized as follows: Section 7.2 describes practical extensions of performance assessment and diagnosis techniques described in chapters 5 and 6. The motivation for proposing these practical methods comes from the fact that the implementation details

of DMC make it quite unique and we need to tailor the analysis tools to suit its needs. The first case study involving C_3F DMC application is described in Section 7.3. A diagnosis of poor performance is attempted using closed loop identification and constraint analysis. Section 7.4 gives the de-methanizer DMC analysis results. The role of LP part of the DMC on performance and stability of the closed loop is highlighted. The role of the lower level PID controllers in diagnosing DMC performance is also emphasized. Section 7.5 gives the concluding remarks.

7.2 Practical Methods for Performance Analysis of DMC

We reviewed and developed some techniques for analyzing the performance of model based predictive controllers in chapters 5 and 6. The application of conventional performance assessment techniques such as minimum variance benchmarking depends on the linearity of the closed loop data. A constrained controller such as DMC does not satisfy these conditions. Also since we are dealing with a commercial control technology here, things like the design objective function are not available. Another practical concern is the multi-rate nature of many DMC applications especially when dealing with measurements of compositions. As we will see in the applications presented in this chapter, compositions are available at a slower sampling rate than the control sampling interval. In the face of these limitations we had to tailor our analysis tools to suit the needs of DMC performance assessment and diagnosis. In the following discussion we present these practical extensions to performance analysis.

7.2.1 Assessment techniques

Multivariable Settling time based measure

Here we extend the performance measure based on the desired closed loop settling time to the MIMO case:

1. Choose a desired settling time, t_c samples. Filter the process outputs (errors) with a simple interactor of the form $z^{-t_c}I$,

$$\tilde{y}(k) = z^{-t_c}y(k).$$

2. The outputs (errors) are pre-whitened using a multivariable auto-regressive (AR)

model,

$$(A_0 + A_1 z^{-1} + ... + A_{nw} z^{-nw})\tilde{y}(k) = e(k) \Leftrightarrow$$
$$(B_0 + B_1 z^{-1} + ... + B_{t_c} z^{-t_c} + ...)e(k) = \tilde{y}(k)$$

where e is a white noise process.

3. Multivariable cross-correlation analysis between the estimated white noise sequence and the interactor filtered outputs gives the desired covariance matrix,

$$\Sigma_{t_c}^2 = (B_0 \sigma_e^2 B_0^T + B_1 \sigma_e^2 B_1^T + ... + B_{t_c-1} \sigma_e^2 B_{t_c-1}^T).$$

4. The multivariable performance index is then obtained as

$$\eta = \frac{tr(\Sigma_{t_c}^2)}{tr(\Sigma_y^2)}. \tag{7.3}$$

In other words $tr(\Sigma_{t_c}^2) = tr(\Sigma_y^2)$ if the achieved settling time under closed loop conditions was equal to the desired settling time. If $tr(\Sigma_{t_c}^2) < tr(\Sigma_y^2)$, then $\eta < 1$ and the achieved settling time is greater than the desired one. The advantage of this approach is that knowledge of the interactor is not needed and the desired settling time requires some a priori information, but is otherwise an easily chosen quantity. Different settling times can be chosen for each output channel in practice. The problem of missing measurements was resolved by assuming a zero-order hold, i.e., the measurements were assumed to be constant at the inter sample instants. This assumption is valid for the case where the process dynamics are sufficiently slow. Theoretically a multi-rate approach to performance assessment is the correct choice.

Achieved Objective function

The (numerical value of) design objective function is not known in practice. In such situations the achieved objective function alone can reveal useful information about the relative changes in performance at every sampling instant. Consider the quadratic form which is present in the objective function of DMC,

$$J_k = \frac{1}{p}\left\{(r_k - y_k)^T \Gamma (r_k - y_k) + \Delta u_k^T \Lambda \Delta u_k\right\} \tag{7.4}$$
$$= \frac{1}{p}\left\{e_k^T \Gamma e_k + \Delta u_k^T \Lambda \Delta u_k\right\} \tag{7.5}$$
$$= \frac{1}{p}\left\{\sum_{i=1}^{P} e_{k+i}^T \Gamma_i e_{k+i} + \Delta u_{k+i-1}^T \Lambda_i \Delta u_{k+i-1}\right\}$$

where p is the prediction horizon, m is the control horizon, the output and the input weightings are given by,

$$\Gamma = diag(\Gamma_1, \Gamma_2, ..., \Gamma_p), \Lambda = diag(\Lambda_1, \Lambda_2, ..., \Lambda_p)$$

and

$$\Delta u_m = \Delta u_{m+1} = ... = \Delta u_p = 0.$$

Also, in practice $\Gamma_1 = ... = \Gamma_p = diag(\gamma_1 ... \gamma_{n_y}), \Lambda_1 = ... = \Lambda_p = diag(\lambda_1 ... \lambda_{n_u})$. Now, with an abuse of notation, if we define $E(.) = \frac{1}{p}\sum_{i=1}^{p}(.)$, then we have

$$J = E(e_k^T Q e_k + \Delta u_k^T R \Delta u_k)$$

where $Q = diag(\gamma_1^2 ... \gamma_n^2), R = diag(\lambda_1^2 ... \lambda_m^2)$. The expectation of the above quadratic form can be expressed in terms of the means and variances of the individual quantities in the following manner.

$$\begin{aligned} J &= E(e_k^T Q e_k + \Delta u_k^T R \Delta u_k) \\ &= m_e^T Q m_e + tr(Q\sigma_e^2) + m_{\Delta u}^T Q m_{\Delta u} + tr(R\sigma_{\Delta u}^2) \end{aligned}$$

where $m_X = E(X), \sigma_X^2 = E\left\{(x - m_X)(x - m_X)^T\right\}$. Thus the achieved objective function is the sum of four quantities - achieved {output offset (m_e) + error variance (σ_e^2) + mean input value ($m_{\Delta u}$) + input variance ($\sigma_{\Delta u}^2$)}. The output offset, output variance and input variance are very important from a control point of view. Monitoring of these quantities on a continuous basis can give insight into the contributing causes of poor performance - offset, input and output variance.

Historical Benchmarking

Often times the controller is known to deliver good performance in a particular region through some performance metric or a priori knowledge such as operator's experience *etc.* We propose a practical index of performance based on comparison of the achieved objective functions in the good region with the current performance. The historical measure of performance is given by

$$\eta = \frac{J_h}{J_a} \tag{7.6}$$

where

$$\begin{aligned} J_h &= E\left\{(r_{k,h} - y_{k,h})^T \Gamma (r_{k,h} - y_{k,h}) + \Delta u_{k,h}^T \Lambda \Delta u_{k,h}\right\} \\ J_a &= E\left\{(r_k - y_k)^T \Gamma (r_k - y_k) + \Delta u_k^T \Lambda \Delta u_k\right\} \end{aligned}$$

r_k, y_k and u_k are the measured values of the setpoints, outputs and inputs respectively. The subscript h denotes the historical data.

7.2.2 Diagnostics

Effect of uncertainty on LP

It is important to consider the effect of gain mismatch on the steady state optimization step in DMC. This step is based on the steady state part (or the gains) alone. Mathematically, a linear programming problem of the following form is solved at every sampling instant in the LP module,

$$\text{Min } \alpha Y_{ss} + \beta \Delta U_{ss}$$

$$\begin{aligned} \text{s. t. } A\Delta U_{ss} + B &\geq 0 \\ Y_{ss} &= K\Delta U_{ss} \end{aligned} \tag{7.7}$$

where K denotes the steady state gains of the step response model. First, we shall illustrate the effect of gain mismatch on the optimum. Consider a 2×2 problem where the true gains between u_1 and y_1, and u_2 and y_2 are higher than the model gains. It is clear from figure 7.3 that, in such cases the estimated optimum may be significantly different from the true optimum depending on the extent of mismatch. In effect, the LP step will be setting the wrong targets. Sensitivity analysis can be used to determine which gains are most important and need to be accurately known.

Valve Related Problems

One of the practical issues that needs to be addressed is how do valve dynamics and non-linearities such as hysteresis, saturation, *etc.* affect performance of the control loop system. From a diagnosis point of view, if a controller is performing poorly when are valve related problems responsible? For some types of valve nonlinearities it is possible to detect the problem by looking at the appropriate input-output plot and detecting *signatures* of these nonlinearities. For a simulation example, figure 7.4 shows the scatter plot of the output against the manipulated variable for valve with significant hysteresis.The discontinuities indicate the presence of a nonlinearity. Butterfly valves are known to exhibit nonlinear behaviour such as hysteresis.

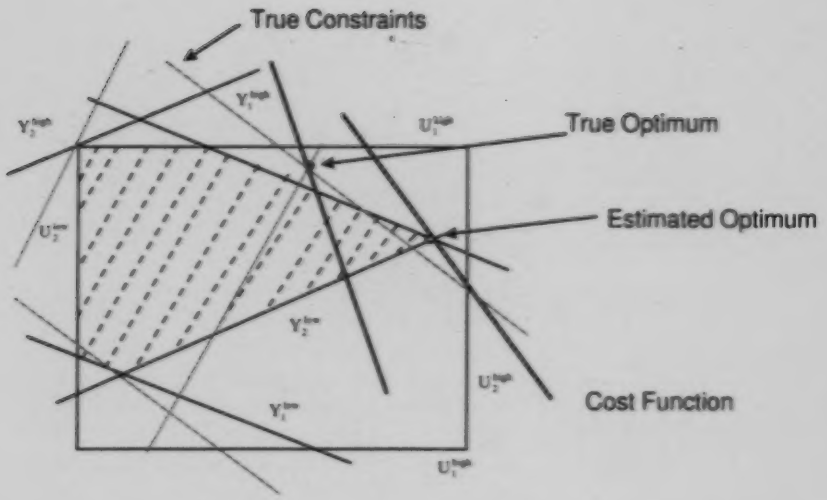


Figure 7.3: *The effect of gain mismatch on the linear programming step in DMC.*

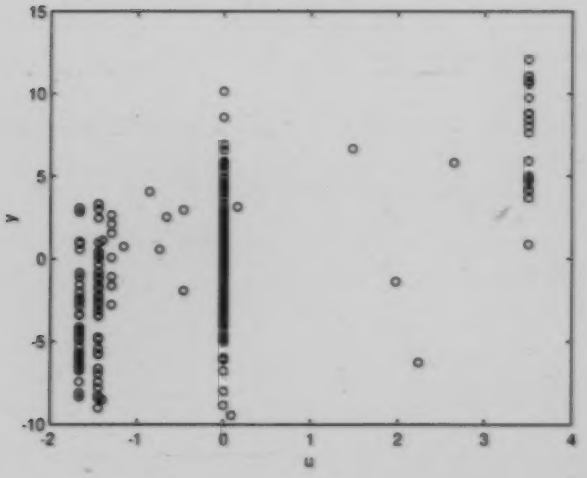


Figure 7.4: *Scatter plot showing a characteristic template for valve hysteresis.*

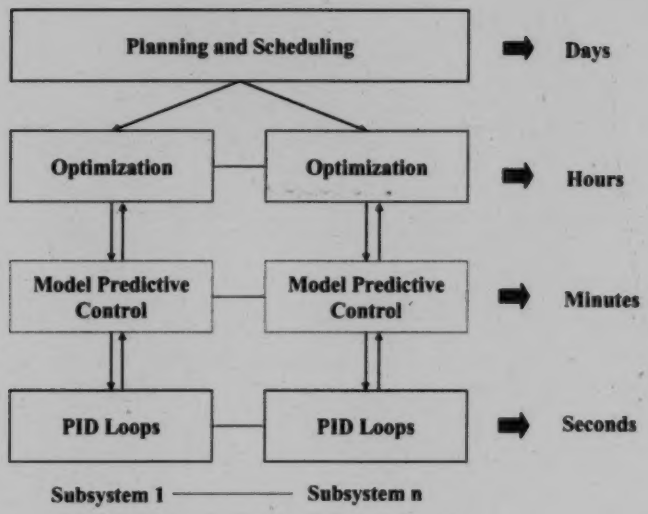


Figure 7.5: *The control heirarchy in a chemical plant.*

Hierarchial Issues

DMC is part of a plant wide control structure illustrated in figure 7.5. At the time of evaluating a higher level control system it is necessary to ensure that all the lower level controllers are functioning satisfactorily. For example a poorly tuned pressure controller in a distillation column can lead to oscillations throughout the column and lead to poor DMC performance. Assessment of lower level controllers can also help in arriving at an early diagnosis of poor DMC performance. The real time optimization layer could set targets for DMC which leads into a highly nonlinear region and thus deteriorate its performance. Thus DMC performance is not independent of the optimization and control layers surrounding it in one or more subsystems. A poorly performing DMC in a neighbouring subsytem will cause disturbances for other DMC applications.

Other factors

Many a time a sensor failure leads to loss of measurement signals and this can be misinterpreted as poor performance. A variety of statistical techniques such as principle component analysis (PCA), PLS, ANN, *etc.* are known to have the kind of fault diagnostic capabilities which can help in detection of sensor malfunction.

Apart from the fundamental limitations on performance discussed in Chapter 6, it should be acknowledged at the outset that DMC is a linear model-based predictive controller

operating on an inherently nonlinear plant. There are limitations which a linear controller cannot overcome for a nonlinear plant no matter how well it is tuned .

In practice, degradation in performance can be caused by a combination of the above factors. It is a challenging research problem to isolate the contribution of individual factors towards poor performance from routine operating data with or without excitation. For the problem under consideration, the propylene splitter DMC, we have resorted to a case study approach. A fair portion of this chapter is devoted to establishing the effect of uncertainty; in order to do this analysis one needs an estimate of the open loop MPM. Closed loop identification is thus necessary to establish changes in plant dynamics.

7.3 Case Study 1: Propylene Splitter DMC

The process under consideration is a propylene splitter (C_3F) column. A schematic of the process is shown in figure 7.6. A dynamic matrix controller (DMC) is currently used to control the key variables of interest which, in this case, are the top and the bottoms impurities. The natural logarithm of the top impurity is controlled as it shows a more linear response. There are a total of 6 controlled variables (CVs), 5 manipulated variables and 2 measured disturbances. The models are available in a step response form obtained from the dynamic matrix identification (DMI) algorithm. The measurements of the main CVs are available at the rate of 20 minutes each.

The controller in place is a multivariable dynamic matrix controller (DMC). The column is operated under normal, low and high load (feed flow rate) conditions depending upon the production requirements (see figure 7.7) . The step response models were originally identified under normal load conditions.

7.3.1 Assessment Results

Performance assessment was carried for different load conditions. Three methods were applied for performance assessment - (1) multivariable settling time benchmarking, (2) historical benchmarking and (3) achieved objective function. The results of the performance assessment step are presented in this chapter.

Normal Load Data

The operating data for the normal load conditions is shown in figure 7.8. The top part of the graph shows the outputs along with the setpoints and the bottom part shows the

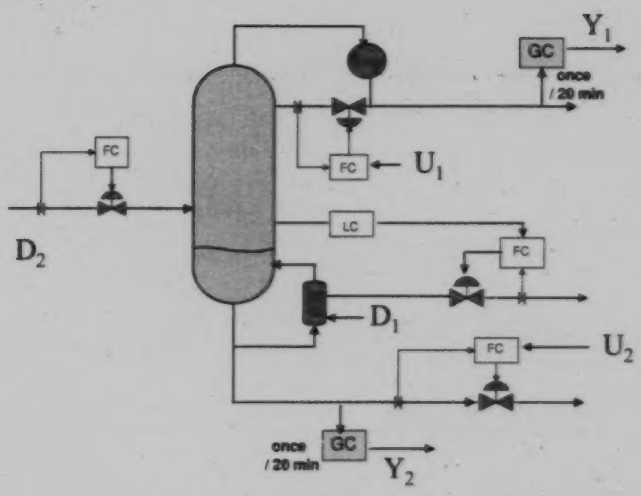


Figure 7.6: *A schematic diagram of the distillation column showing the relevant controlled and manipulated variables.*

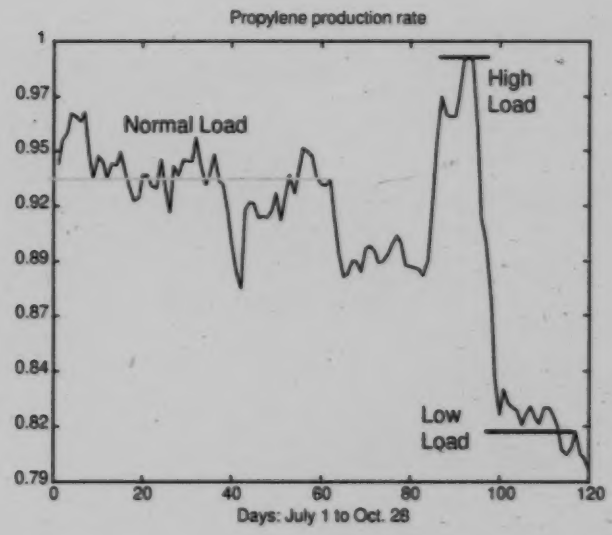


Figure 7.7: *The Production requirements from July 1 to Oct. 28. The data sets corresponded to the highlighted regions.*

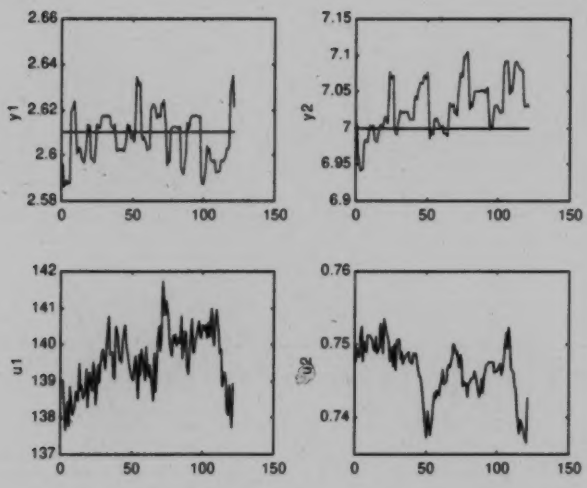


Figure 7.8: *The input-output data for the normal load conditions*

manipulated variables. The sampling time for control is 6 minutes and the data shown here corresponds to each control sampling interval.

A desired settling time of 5 samples was chosen for each of the outputs based on the open loop settling time and the recommendations of the plant personnel. Since the settling time index was (0.86) high for this data set we chose the normal load data as our benchmark data for historical benchmarking. Figure 7.9 shows the performance indices for the cases where (1) the weightings were taken into account and (2) the raw data was used. For this DMC application the output weighting matrix is given by:

$$\Gamma = diag(200, 2).$$

The top impurity is the economically important variable and is weighted by a factor of 100 times more than the bottom. It is important to consider these weightings in any benchmarking scheme.

For the normal data set the settling time measures showed good performance irrespective of the weightings. The historical benchmark measure equals 1 since this data set itself is the benchmark data set. The achieved objective function could not be calculated due to insufficient data. For the remaining data sets we do not show the data trends but the assessment results are summarized below.

The results for the high load data are reported below. The settling time and the his-

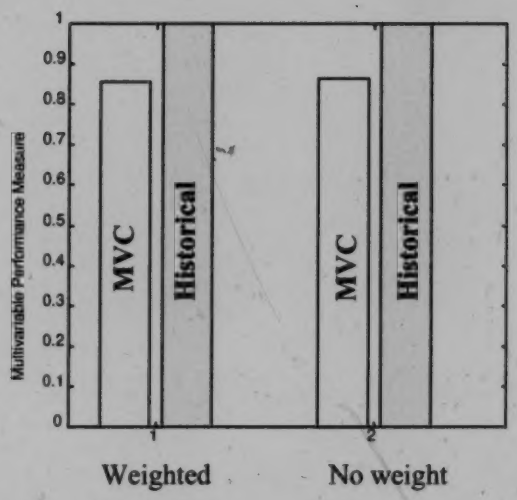


Figure 7.9: *The multivariable minimum variance performance index and the historical benchmark, for the (i) raw and (ii) weighted data.*

torical measures are shown in figure 7.11. Both the measures indicate a significant drop in performance compared to the normal load conditions. In the later Sections we will attempt to isolate the cause for this drop in performance. The achieved objective function for the last 60 samples is shown in figure 7.10. The achieved objective function shows a decreasing trend, which implies improving performance. The output variance term is dominant compared to the offset and the input energy terms. For the first low load data set, both the settling time and the historical measures indicate a drop in performance for the low load case. For this case the achieved objective function was dominated by the offset term. The contribution of the input variance term is negligible.

The overall performance assessment results are shown in figure 7.11. The settling time approach indicates good performance in the normal load region, poor performance in the high load region. The first low load data set showed poor performance. The constraints were active throughout this run. The second low load data set showed satisfactory performance. The constraints were active only for the initial part of this run (see figure 7.12).

The historical benchmarking method confirmed the settling time approach based results except for the second low load data set. For this data, the historical benchmark indicated no significant improvement in performance. This can be attributed to an increase in input variance or output offset, since the historical benchmark takes them into account explicitly.

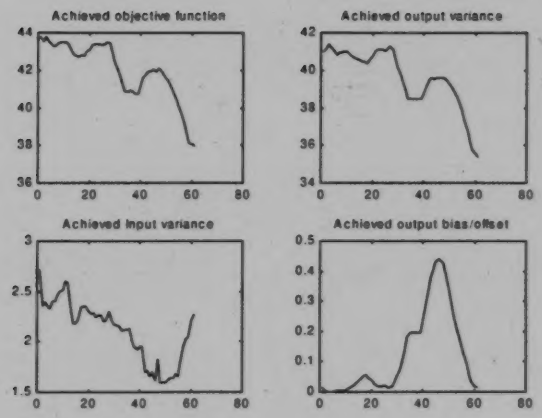


Figure 7.10: *The achieved objective and the contributing factors for the high load data.*

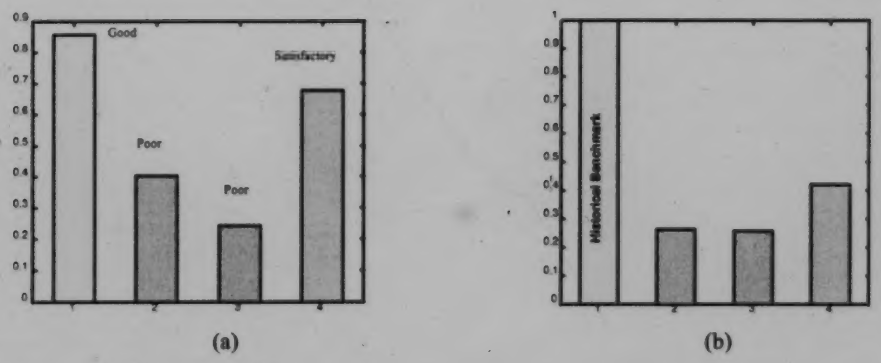


Figure 7.11: *(a) The MVC performance indices for different data sets (1-Normal, 2-High, 3-Low A, 4-Low-B) and (b) the historical performance measures.*

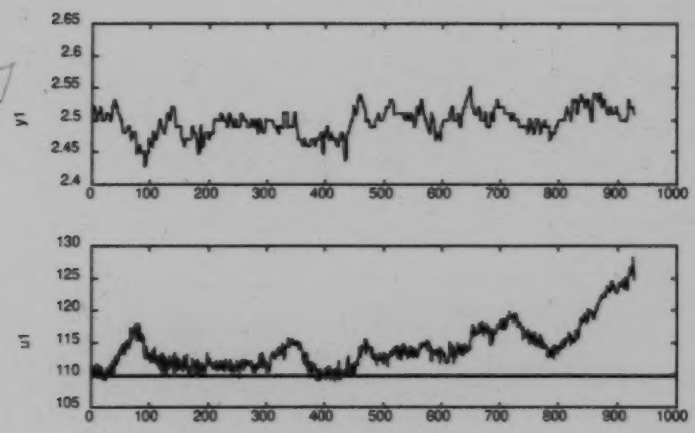


Figure 7.12: *The input-output data for the low load (B) data set. The reflux flow rate is at the lower operating limit.*

7.3.2 Diagnostics

Analysis of constraints

During the low load conditions the lower constraint on reflux flow rate was activated. Material balance considerations can be used to explain this phenomenon. The lower limit on the input constraint was an operator specified quantity. Since the reflux flow rate was at the constraint, the control system was effectively reduced to a single input two output system. Theoretically this is an uncontrollable system. Economically, the top impurity is more important than the bottom impurity, and due to the emphasis placed on this output, the bottom flow rate was devoted entirely to controlling output 1 during this period. An offset was observed in output 2 for a part of the low load region.

Prediction Errors in the closed loop

In order to investigate the reasons for poor performance we investigated the prediction errors for the DMC controller. DMC uses a step response model of the process combined with an estimated disturbance to predict the process outputs over the prediction horizon. The disturbance is estimated through feedback:

$$\hat{d}(k/k-1) = y(k-1) - \hat{y}(k-1)$$

where y is the measured value and $\hat{y}$ is the predicted output.

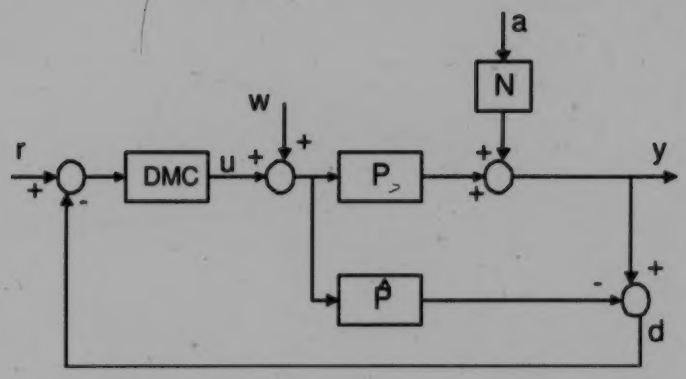


Figure 7.13: *The internal model control structure for DMC*

The open loop predictions are updated with this disturbance estimate. This step in DMC allows us to rewrite the DMC controller in the IMC form. For the propylene splitter DMC, the step response model is also used to predict the compositions between two measurements. This is equivalent to an observer in control terminology. From figure 7.13 we can write down the prediction errors in the closed loop as,

$$d = \Delta(I + Q\Delta)^{-1}w + N(I + Q\Delta)^{-1}a + Q\Delta(I + Q\Delta)^{-1}r \tag{7.8}$$

where Δ is the uncertainty, a is disturbance, r is the setpoint and w is the dither signal or external excitation. If $w = 0$, we have,

$$d = N(I + Q\Delta)^{-1}a + Q\Delta(I + Q\Delta)^{-1}r = (I + \Delta Q)^{-1}\{\Delta Qr + Na\}.$$

Thus, the prediction errors, under closed loop, are influenced by the controller, Q and the extent of uncertainty, . However, notice that if there is no MPM and no external signals (r and a), $d = 0$ irrespective of the controller. This observation forms the basis for deciding if the model is satisfactory. If several factors, other than mismatch, could be held responsible - Q, N, r, a.

In an effort to isolate the reasons of poor performance we looked at the infinite horizon prediction errors of the step response model under closed loop conditions. The following figures (7.14 and 7.15) shows the prediction errors for the low and high load conditions respectively.The model seemed to worsen considerably under high load. It should be noted here that we are dealing with closed loop data. Hence, the controller and/or unmeasured disturbances can affect the predictive capabilities significantly, under closed loop. Kesavan

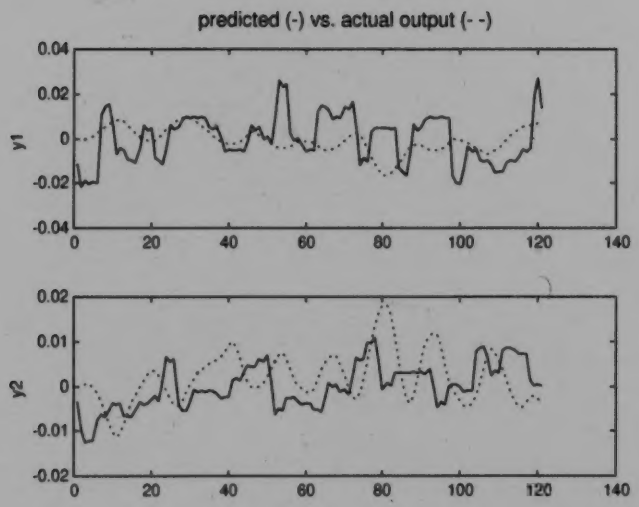


Figure 7.14: *Prediction errors for the DMI model: low load.*

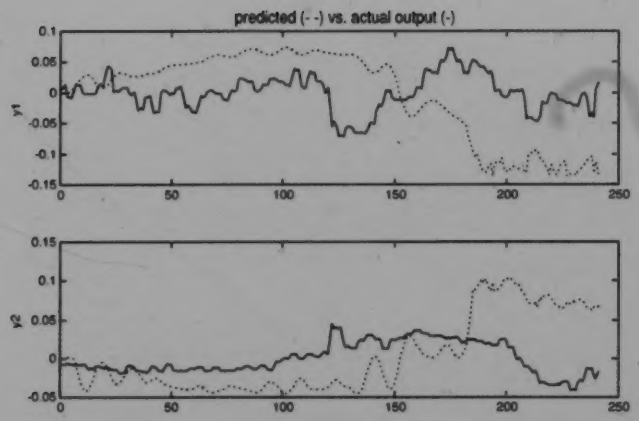


Figure 7.15: *Prediction errors for the DMI model: high load.*

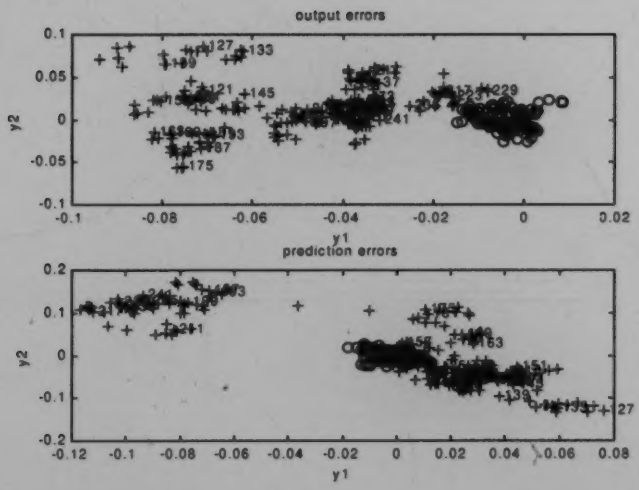


Figure 7.16: *The top graph shows the output errors of y2 vs. y1 and the bottom graph shows the corresponding prediction errors. The circles (o) denote the data from the low load conditions while the crosses (+) denotes high load data. Every sixth point from the high load data set is numbered.*

and Lee (1997) also suggested looking at the prediction errors for diagnosis of poor performance. Their approach consists of looking at the prediction errors in conjunction with the output errors or deviations from the setpoints. Figure 7.16 shows these trends jointly for both the outputs.

Figure 7.16 shows that the predictions are considerably worse for the high load conditions and the predictions have deteriorated in two stages - set 1 (13-181) and set 2 (187-241). However the output errors are decreasing for set 2 (see top graph in figure 7.16). Thus it is useful to look at these two quantities together. Note that the measured disturbance, feed flow rate, was part of the prediction model.

Low Load Data Set 1

The prediction errors for the low load data are shown in figures 7.18 and 7.19. For the second low load data set the predictions for output 2 appear to have worsened considerably.

Quantification of prediction errors is important in establishing model quality. Kozub (1997) proposed the use of minimum variance benchmarking on the prediction errors to obtain a prediction index. We applied this idea for the C_3F splitter DMC with the following results.

This measure can be treated as a necessary test for goodness of model quality. If it equals unity, then the model quality is good. If it is less than unity, several reasons can be

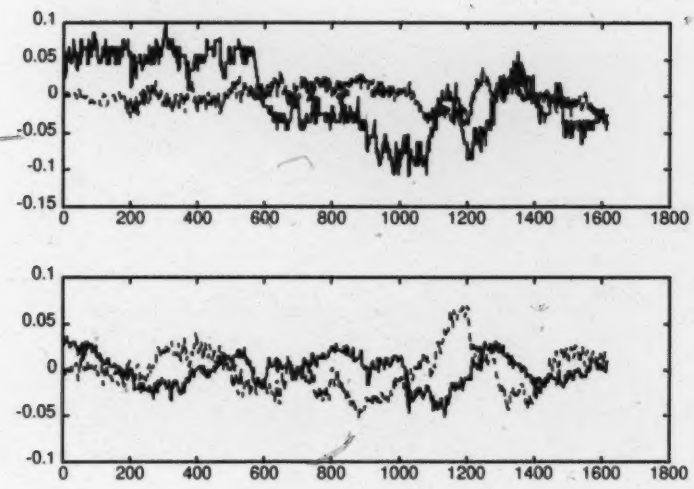


Figure 7.17: *The prediction errors for low load data set A.*

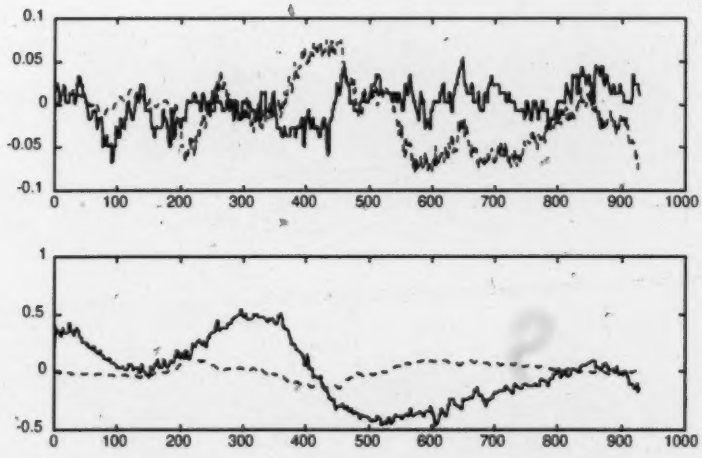


Figure 7.18: *The prediction errors for low load data set B.*

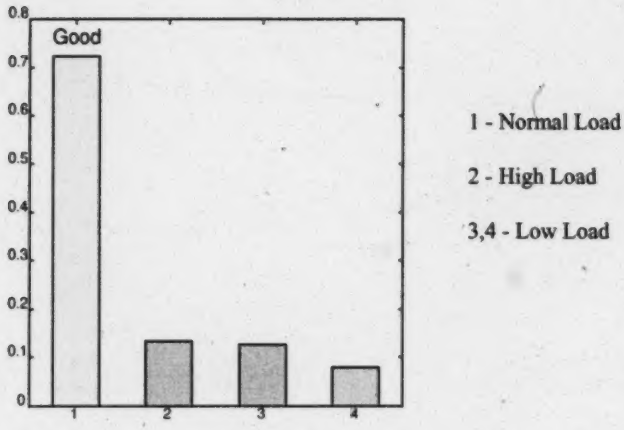


Figure 7.19: *Measures for prediction errors based on the minimum variance benchmark.*

listed for its deviation from unity and no statement can be made about the model quality. For our data sets, the MVC measure indicated satisfactory model quality for the normal load case relative to the other data sets.

Identification in the Closed Loop

It is of great practical and theoretical interest to estimate the model parameters under closed loop to evaluate the change in plant dynamics. Many different approaches to closed loop identification have been proposed in literature to estimate the open loop model parameters. These include - direct identification, indirect identification and joint input-output identification. It is beyond the scope of this report to discuss these methods in detail. The interested reader is referred to Van den Hof and Schrama (1995) for a survey of the closed identification methods.

Direct closed loop identification involves the direct use of input-output data in closed loop and an appropriate noise model. It is necessary to remove the effects of feedback when using the input-output data directly. A high signal-to-noise ratio and/or changes in the controller are necessary to prevent the controller inverse being identified. For our data sets, surprisingly, single variable correlation analysis (see the *cra* function in Matlab) yielded the best results,

$$y(k) = \sum_{i=1}^{N} h_i u(k-i) + e(k). \tag{7.9}$$

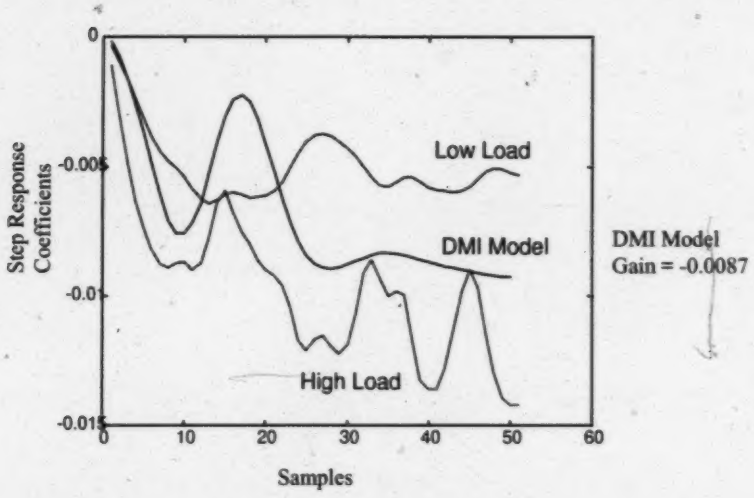


Figure 7.20: *The finite impulse response models between the top impurity and the reflux flow rate. These models were estimated for the low and high load conditions using univariate correlation analysis (after prewhitening) and are compared with DMI model.*

The results for single variable correlation analysis (finite impulse response models) are shown below for the relationship between top impurity and the reflux flow rate.

The estimates during the high load conditions are noisy due to the small number of samples (241) available for estimation. From the general trend, it can be inferred that the gain increases during the high load conditions and decreases during low load conditions. This was verified through simulation using a steady state model of the process. The magnitude of the gain changes, however does not appear to be significant.

Chapter 7 addresses the issues of identifiability under closed loop conditions using correlation analysis. The second observation, while using correlation analysis, was the effect of the pre-whitening filter on the estimation process. Shown below are output error (OE) models between the top impurity and the reflux flow rate. When the OE structure is used without pre-whitening, a positive gain is identified and dynamic response is entirely different from the open loop model. When a pre-whitening filter is used to whiten the inputs, and the output is filtered using the same filter, the estimated model is much closer to the

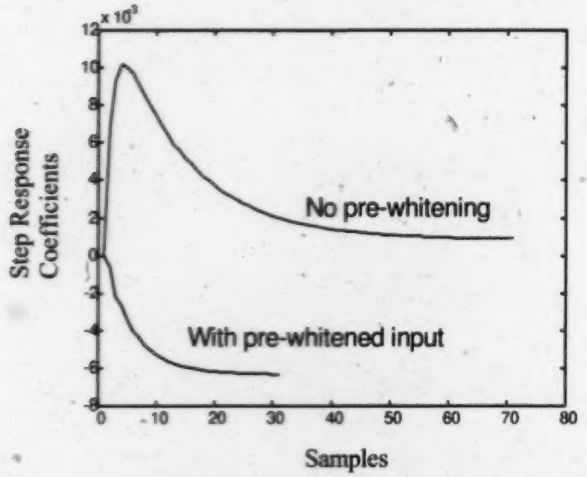


Figure 7.21: *Output error (OE) models between the top impurity and the reflux flow rate with and without pre-whitening. The effect of feedback is clearly seen when no pre-whitening is done.*

open loop characteristics,

$$\begin{aligned} y(k) &= \sum_{i=1}^{N} h_i u(k-i) + e(k) = H(z^{-1})u(k) + e(k) \qquad (7.10) \\ L(z^{-1})y(k) &= H(z^{-1})L(z^{-1})u(k) + L(z^{-1})e(k) \\ y_f(k) &= H(z^{-1})a(k) + e_f(k) \end{aligned}$$

where L is the pre-whitening filter. This phenomenon is explained in Chapter 8 in more detail.

Discussion

The statistics for assessment of control errors and the prediction errors are reported in Table 7.1. We have resorted to a case study approach for the diagnosis of poor performance, in the high and low load regions. The main results of the performance diagnosis phase are summarized in Table 7.2. The different possible causes of poor performance are ranked in order of importance (1 is highest). Currently this ranking is done qualitatively. Quantifying the contribution of different factors is a challenging research problem. It should be noted that the different factors are often correlated. For example, during the high load conditions there was a disturbance in the feed flow rate which caused a change in plant dynamics and the controller tunings could be inappropriate due to these changes.

Constraints were shown to play an important role in the analysis of performance. Some evidence of MPM was found through correlation analysis.

7.4 Case Study 2: De-methanizer DMC3

The DMC application, on the de-methanizer Section of the plant, consists of 18 CVs and 15 MVs. The dynamic matrix has 87 non-zeros (30%). In the subsequent discussion we focus on a 6x6 subsystem. The main CVs of interest are top and bottom impurities, column pressure, a tray temperature and the reboiler level. The three compositions are sampled at 30 minute intervals, the control sampling interval is 4 minutes. The output weighting matrix is given by

$$\Gamma = diag(0.0133, 5.556, 0.10, 10.00, 1.00, 0.10).$$

At the outset, we would like to mention that this appeared to be a highly de-tuned application with most of the move suppression factors being of the order of 100.

7.4.1 Assessment Results

Here we give the performance metrics based on the desired settling time idea. The data was divided into two batches and the performance was assessed for desired sampling times of 7 and 8 samples (28 and 32 minutes) respectively. The choice of the desired settling time was based on the sampling interval for the 3 compositions which was 30 minutes. The multivariate performance indices are shown in figure 7.22. The relative performance for the second batch of data is poor. Since the data is weighed heavily by the second and the fourth outputs, we also looked at the univariate measure for these two outputs separately.

Figures 7.23 and 7.24 show the individual performance indices for output 2 and output 4 respectively. Clearly the multivariable performance index shows trends similar to that of CV 4. The data for the second batch shows a large upset. It was found that the DMC was taken offline for a short period between sampling instants 3250 and 3375 for some maintenance work and this indeed reflects in the multivariate and univariate performance measures.

In addition a large offset was observed in the pre-fractionator bottoms impurity shown in figure 7.25.

Only the reboiler flow rate has any effect on the pre-fractionator bottoms impurity. The step response model between these variables indicated a low gain (-0.0141). This coupled with the fact that the move suppression factor for the reboiler flow rate is very large - 200

Table 7.1: *Performance measures and the prediction measures for the propylene splitter DMC application.*

	Normal Load	High Load	Low Load 1	Load Load 2
Performance Measures				
Settling time Benchmark	0.8572	0.4026	0.2423	0.6790
Historical Benchmark	1.000	0.2637	0.2544	0.4181
Achieved Objective	-	42.24	105.07	25.2
Constraints				
Input Constraints			Active	Active
Output Constraints	Violations	Violations		
Prediction Measures				
Settling time based index	0.6841	0.1440	0.1196	0.1647
Prediction error index (Historical)	1.000	0.0254	0.0620	0.0803
Overall Peformance	Good	Poor	Poor	Satisfactory

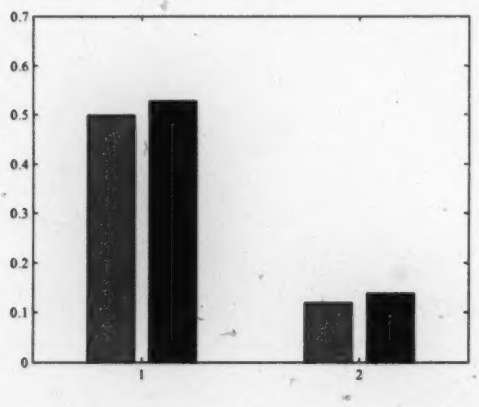


Figure 7.22: *Multivariate performance measures for the demethanizer DMC based on the settling time index.*

Table 7.2: *Summary of performance diagnosis results for the propylene splitter DMC application.*

	High	Low 1	Low 2	Remarks
Diagnostics				
Model plant mismatch	1	2	3	The extent of MPM was assessed through closed loop identification
Constraints	-	1	2	These indices consider both
Poor Tuning	3	3	4	Poor tuning is often related to factors 1 and 2.
Sensor Failures	-	-	-	No sensor failures were detected
Disturbance-Unmeasured	-	-	-	No unmeasured disturbances were detected
Disturbance-Measured	2	4	1	In this case the main measured disturbance is the feed flow rate which in turn leads to MPM and activation of constraints
Confidence Level	high	high	high	
Overall Peformance	Poor	Poor	Satisfactory	

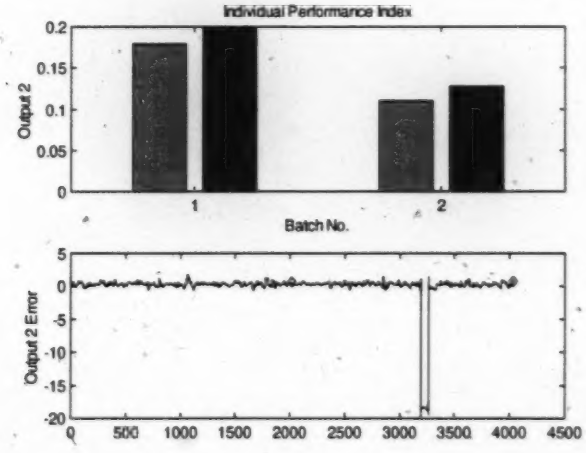


Figure 7.23: *Individual performance measures for output 2, along with the output control error trends.*

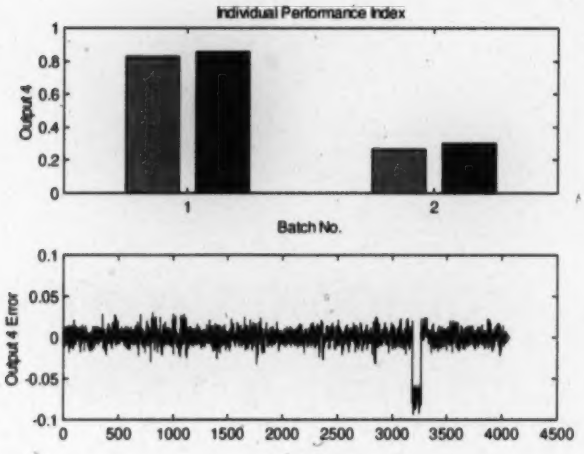


Figure 7.24: *The univariate performance index for output 4.*

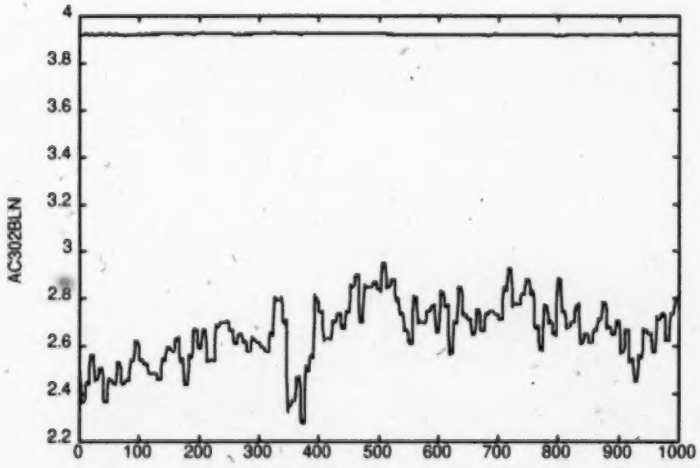


Figure 7.25: *The pre-fractionator bottoms impurity shown along with the setpoint.*

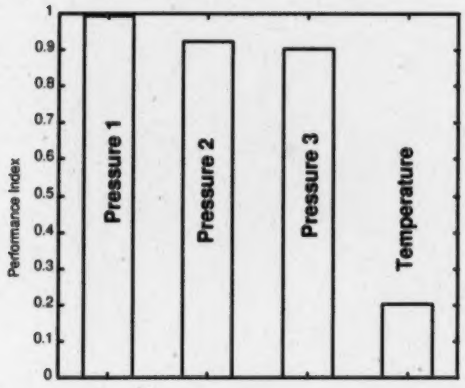


Figure 7.26: *Performance assessment for the PID loops.*

- means that there is a controllability problem with this bottoms impurity and no other input can move it to the desired value.

7.4.2 Diagnostics

PID tuning

As part of the DMC performance assessment, the lower level PID loops were assessed for possible poor performance. The performance assessment results based on desired settling time are shown in figure 7.26. The desired settling time was chosen as 4 minutes, since that is the DMC sampling interval, and the PID dynamics should not be seen at the DMC stage. According to this measure, a temperature loop was found to be poorly performing. The remaining pressure loops showed satisfactory performance.In order to diagnose the temperature loop, closed loop identification was carried with the setpoint as the input and the temperature as the output. It should be noted here that the setpoint to this loop is a manipulated variable for DMC and thus there is sufficient excitation available to carry out closed loop identification. A second order output error model was identified. Figure 7.27 shows the closed loop pole locations for the temperature loop. The poles show very poor damping properties. Also surprising is the presence of a non-minimum phase zero since this cannot be explained in terms of the physical understanding of the process. The interaction of the temperature loop with the DMC application on top is suspected to be the reason for this inverse response in temperature. The closed loop step responses of the temperature loop are shown in figure 7.28. The transient response shows the poor damping properties in time domain, the closed loop settling time is of the order of 90-100 minutes.

On the time scale of the DMC application this is equivalent to 23-25 samples which is not desirable.In order to improve the performance of the PID controller an open loop test was

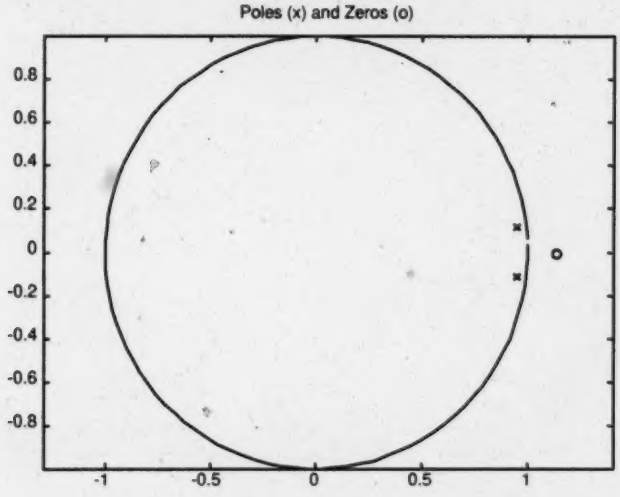


Figure 7.27: *Closed loop pole locations for the temperature loop.*

conducted. The open loop test revealed severe nonlinearities which included variable time delay, changing dynamics (heating & cooling) and an integrating response. Re-tuning of the PID controllers was attempted based on this open loop test but it did not lead to improved performance. The severe nonlinearity of the temperature process was believed to limit the performance of a PID controller and an advanced control scheme is needed to compensate the inherent nonlinearity of the process.

The LP targets

The setpoints (LP targets) for the column pressure and the reboiler level are shown in figure 7.29. For this data set, the LP targets or the setpoints showed greater variability than the outputs. This is a very surprising fact since the DMC controller is ideally designed to track step changes in setpoint whereas in this case it is being made to track an oscillating signal. In fact the setpoints or the LP targets to the controlled variables looked more like a stochastic input than a regular setpoint. In an effort to understand this phenomenon, we looked at the DMC structure in detail.

The combined structure of DMC is illustrated in figure 7.30. The main steps in DMC can be explained in the following manner.

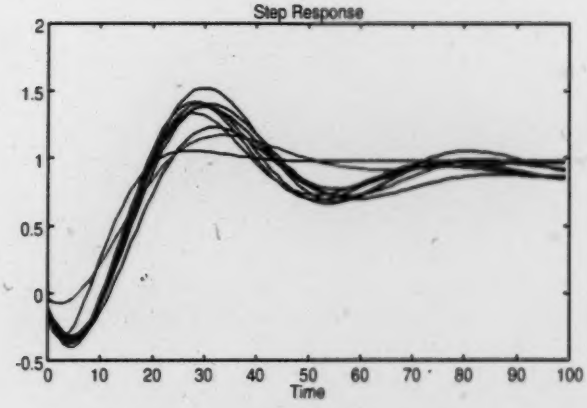


Figure 7.28: *The closed loop step responses for a period of time spanning over 11 days.*

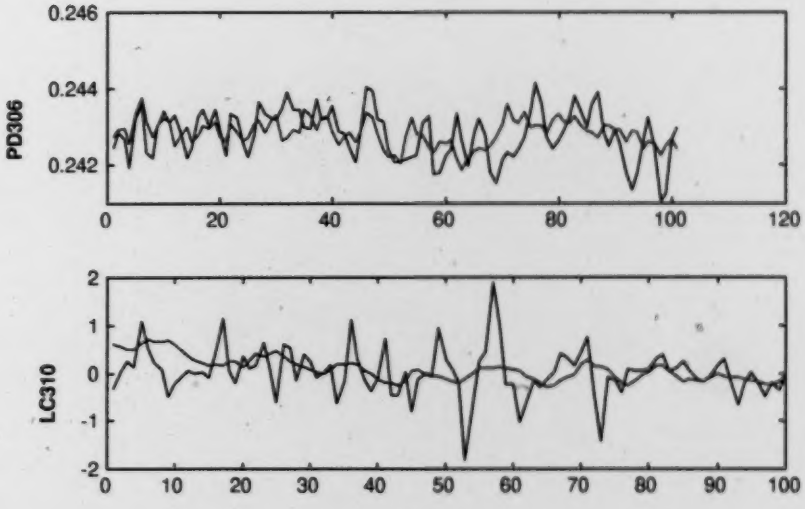


Figure 7.29: *The LP targets along with the outputs for the column pressure and the reboiler level.*

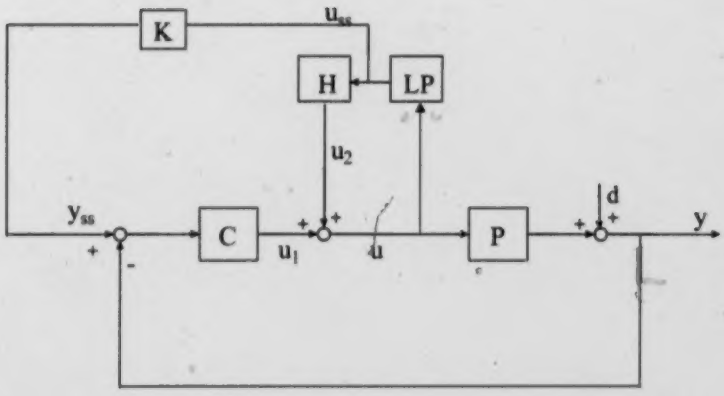


Figure 7.30: *Interpretation of the DMC structure including the LP step.*

1. U_1 is the control move calculated as a result of the unconstrained least squares minimization. C is a transfer function that can be derived on the basis of the least squares solution (see Chapter 2). This unconstrained move may violate one or more constraints on the inputs\outputs.

2. The LP step uses the previous implemented input, $u(k-1)$, and the measured output, $y(k-1)$, to set up the constraints and obtain targets for the inputs through steady state optimization - u_{ss}. This is a nonlinear step, since different constraints may be active at any given instant.

3. H is a transfer function closely related to C, and can be calculated from the unconstrained least squares solution. u_2 $(= Hu_{ss})$ is used to correct the unconstrained control move u_1 and make it conform to the constraints in the least squares sense.

4. The steady state targets for the controlled variable (setpoints) are obtained by multiplying the input steady state targets with the gains, K.

The main outcome of the LP step choosing the setpoints is that the feedback decides setpoints, and variability is transferred from the input/output to the setpoint. From a control design view point, the linear programming step is really a part of the feedback control system and any stability and/or performance analysis would have to include the LP part. The design of the LP cannot be performed independent of the dynamic error minimization step. The stability of the dynamic error minimization step does not guarantee the stability or performance of the overall system.

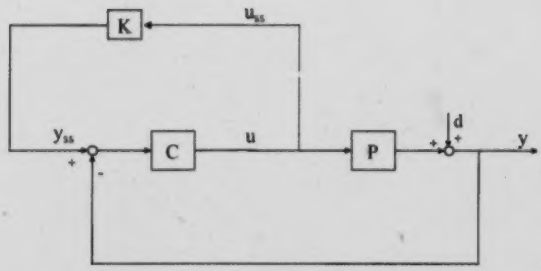


Figure 7.31: *The feedback representaion of DMC LP step.*

Ideally the setpoint changes should be implemented at a rate which matches the closed loop settling time of the control system. A low pass filter on the setpoint could be used to achieve this effect.

Stability analysis of the DMC controller

Clearly the preceding discussion indicated that the LP step affects the stability of the closed loop system. In the ensuing discussion we will try to analyze the closed loop stability analysis of DMC including the LP step. The LP step is in reality a nonlinear operator the output of which depends on the set of active constraints and the input data at any given sampling instant. This, in turn, makes the overall closed loop system nonlinear and the stability analysis is rather difficult. Here, we treat the LP step as an uncertain gain element (bounded) and then invoke classical robust stability analysis results, the small gain theorem, to derive sufficient conditions for closed stability of the combined DMC system.

Theorem 7.1. *For the feedback system in the figure 7.31 , the closed loop system is robustly stable if and only if*

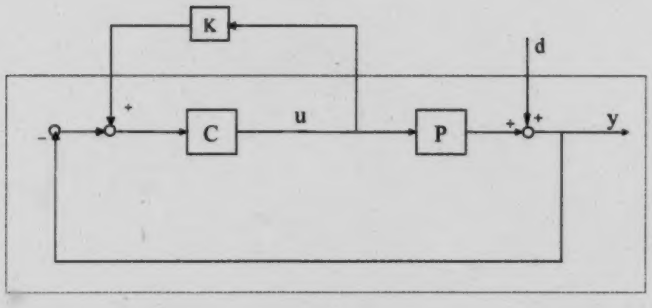
$$\|C(I+PC)^{-1}K\|_\infty < 1 \Leftarrow \tag{7.11}$$
$$\|C(I+PC)^{-1}\|_\infty < \frac{1}{\|K\|_\infty}$$

For a SISO system with positive K we have the following condition for robust stability

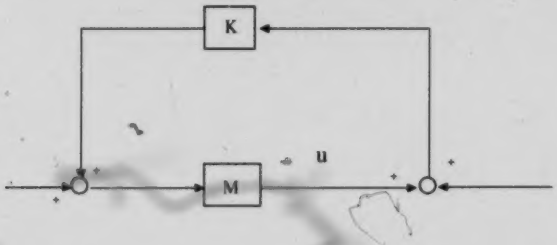
$$\|\frac{C}{1+PC}\|_\infty < \frac{1}{K}$$

Proof. Consider the feedback system in figure below

where K is an uncertain gain matrix depending upon the set of active constraints. The above figure is equivalent to



This can then be expressed in the classical M-Δ structure for feedback uncertainty



with $M = C(I + PC)^{-1}$. Therefore using the small gain theorem we have for robust stability

$$\begin{aligned} ||M\Delta||_\infty &< 1 \Leftrightarrow \\ ||C(I+PC)^{-1}K||_\infty &< 1 \end{aligned}$$

Q.E.D ∎

Corollary 7.1. *For a SISO system, such that $\frac{C}{1+PC}$ has a monotonically decreasing magnitude response and C has an integrator (pole at z=1), the robust stability condition reduces to*

$$K < P(1)$$

where P(1) is the process gain.

7.5 Conclusions

The main objectives of this work were to assess the performance of the C3F splitter DMC and subsequently analyze the reasons for poor performance. At the assessment stage, performance was found to be poor in the low and high load operating regions. Settling time (Kozub and Garcia, 1993) and historical benchmarks were found to be useful in quantifying

DMC performance for this application. The achieved objective function is useful in quantifying the different contributions towards poor performance and is recommended for future online use.

At the analysis stage, constraints were found to be the main reason for the degradation of poor performance. When the column went into the low load region, material balance considerations led to the activation of the lower constraint on the reflux flow rate. The controller was effectively reduced to a single input two output system in this period. In the high load region a combination of measured and unmeasured disturbances were responsible for loss in performance.

The chief conclusions of this case study are as follows:

1. Even advanced controllers like DMC may be delivering poor performance. Univariate and multivariate controller performance assessment is necessary to monitor these loops regularly and detect performance degradation.

2. It is necessary to document the design performance of advanced and lower level controllers. This would be extremely useful in performance assessment as well as diagnosis. A controller database which documents the basic process characteristics such as time delay, time constant, gain, and the design expectations such as designed closed loop response, *etc.*, would go a long way in developing consistent performance measures and diagnostic tools.

3. Diagnosis of performance was performed on a case study basis. Automated performance diagnosis and subsequent remedial action is a long term goal of this work. The experience gained during this work will prove useful in establishing a performance analysis framework.

4. Constraints were shown to play an important role in the diagnostics step. The interaction of the constraints and the induced variability in the LP targets or the setpoints was found to be related to implementation issues in DMC.

5. Closed loop identification has great potential in establishing the causes of poor performance. Effective multivariable closed loop identification methods need to be developed for use with large dimensional industrial data sets.

Chapter 8

Towards Iterative Control and Identification for MPC

8.1 Introduction

The area of control relevant identification or joint identification and control focuses on the need to have matching identification and control criterion in order to design high-performance controllers. In conventional system identification the focus has been on consistency and estimation of the true plant from the measured data. In practice, bias (undermodelling) and variance errors will always be present in the estimated model. Control relevant identification realizes the need to put emphasis on shaping the bias and variance errors so as to minimize their impact on closed loop performance of the control system. In the following discussion we give an outline of the area of control relevant identification as per Van den Hof and Schrama (1995). In model-based control a controller C is designed based on a nominal plant, $\hat{P}$, based on some design objective denoted by $J(\hat{P}, C)$. This controller is then implemented on the true plant P. The controller will robustly stabilize the true plant only if

$$\left\| (P - \hat{P})C(I + \hat{P}C)^{-1} \right\|_{\infty} < 1 \tag{8.1}$$

based on the small gain theorem. The achieved performance will in general be different than the designed performance. For example the degradation in tracking performance is given by the difference in the achieved and designed complementary sensitivity functions,

$$\begin{aligned} & PC(I + PC)^{-1} - \hat{P}C(I + \hat{P}C)^{-1} \\ = \; & (I + PC)^{-1}(P - \hat{P})C(I + \hat{P}C)^{-1}. \end{aligned} \tag{8.2}$$

[0] A version of this chapter has been submitted for possible presentation at the 1999 AIChE annual meeting in Dallas, TX., as "Control Relevant Identification for Model Predicitive Control: Some Closed Loop Issues", R. S. Patwardhan, R. Gopaluni and S. L. Shah.

If 8.2 is small then robust performance is guaranteed. In the IMC formulation the robust stability and robust performance conditions respectively reduce to

$$\begin{aligned} \|Q\Delta\|_\infty &< 1 \\ \left\|\hat{S}[(I+Q\Delta)^{-1}-I]\right\|_\infty &< \epsilon. \end{aligned} \tag{8.3}$$

The design criterion $J(\hat{P},C)$ depends on the design philosophy adopted for controller design. Many design methods are based on minimization of a single measure of performance, e.g.:

1. *LQ Tracking:* The LQ criterion,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}E\{[r(k)-y(k)]^2+\lambda u^2(k)\} \tag{8.4}$$

is equivalent to $||J(P,C)||_2^2$ with

$$J(P,C)=[r(k)-y(k)\ \lambda u(k)]^T$$

assumed to be a quasi-stationary signal[2].

2. *Mixed sensitivity optimization*: The design objective is indicated by the frequency domain terms,

$$J(P,C)=\begin{bmatrix} W_1(I+PC)^{-1} \\ W_2PC(I+PC)^{-1} \end{bmatrix} \tag{8.5}$$

with the weighting function W_1, W_2 and the corresponding cost function given by $||J(P,C)||_\infty$.

3. *Model predictive control:* The MPC objective function,

$$\sum_{i=1}^{p}||r(k+i)-y(k+i)||_{\Gamma_i}^2+\sum_{i=1}^{m}||\Delta u(k+i-1)||_{\Lambda_i}^2$$

is equivalent to $||J(P,C)||_2^2$ with

$$J(P,C)=[\Gamma^{1/2}(r_k-y_k)\quad \Lambda^{1/2}(\Delta u_k)]^T \tag{8.6}$$

where $||.||_2$ is the conventional Euclidean norm of a vector, r_k, y_k and Δu_k are vectors of appropriate dimensions defined in Chapter 2 and $\Gamma = diag(\Gamma_1,...,\Gamma_p), \Lambda = diag(\Lambda_1,...,\Lambda_m)$.

[2] $||.||_2$ is used to denote the correponding semi-norm, defined by $||x||_2^2 \triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}E[x^T(k)x(k)]$.

The survey by Gevers (1993) gives an excellent justification of the need for joint design of identification and control. The objectives of control relevant identification are illustrated below (Van den Hof and Schrama, 1995)

$$\Big|\,||J(\hat{P},C)|| - ||J(P,C) - J(\hat{P},C)||\,\Big| \leq ||J(P,C)|| \leq ||J(\hat{P},C)|| + ||J(P,C) - J(\hat{P},C)||. \tag{8.7}$$

In this triangle inequality,

$||J(P,C)||$ is the achieved performance,

$||J(\hat{P},C)||$ is the designed performance,

$||J(P,C) - J(\hat{P},C)||$ is the performance degradation.

Clearly if the performance degradation is small, the achieved performance will be close to design performance. This is the robust performance condition. Iterative schemes based on control relevant identification use the past performance degradation term as the identification objective to estimate the model parameters,

$$\hat{P}_{i+1} = \min_{\hat{P}} ||J(P,C_i) - J(\hat{P},C_i)||. \tag{8.8}$$

Thus the identification criterion is commensurate with the control criterion and will decide what errors are relatively more important from the controller's point of view.

The identified model is then used for controller design subject to the constraint that the future performance degradation term, which is function of the newly estimated model and the to be designed controller, is significantly less than the designed objective,

$$\begin{aligned} C_{i+1} &= \min_{\tilde{C}} ||J(\hat{P}_{i+1},\tilde{C})|| \\ &\quad \text{subject to} \\ ||J(P,\tilde{C}) - J(\hat{P}_{i+1},\tilde{C})|| &\ll ||J(\hat{P}_{i+1},\tilde{C})||. \end{aligned} \tag{8.9}$$

It is necessary to have an uncertainty description in order to ensure the satisfaction of the robust performance constraint. In general the controller may not satisfy this constraint. In such cases it is necessary to either choose a higher complexity controller or re-parametrize the model in order to reduce the uncertainty. The iterative scheme is implemented in semi-adaptive mode. Zang *et al.* (1995) gave an iterative scheme based on the LQ objective function. The LQ-relevant performance degradation term was chosen as the identification objective. The following pre-filter when used with the conventional identification objective function led to minimization of the LQ-relevant identification criterion,

$$|L_z|^2 = (1 + \lambda|C|^2)|S|^2. \tag{8.10}$$

where S is the sensitivity function. Lee *et al.* (1993) proposed an iterative identification scheme based upon the internal model control paradigm. This approach is referred to as the *windsurfer* approach in the literature. The IMC performance criterion is chosen as

$$\|J_{IMC}(P,C)\|_\infty = \left\| \frac{PC}{1+PC} - T_d \right\|_\infty \tag{8.11}$$

where T_d is some desired complementary sensitivity function. This performance criterion induces the identification criterion

$$\left\| J_{IMC}(P,C) - J_{IMC}(\hat{P},C) \right\|_\infty .$$

Schrama and co-workers used a criterion similar to the mixed sensitivity function in 8.5 except that their criterion is more general and incorporates disturbance rejection, sensitivity and robustness margins. Their performance objective can be considered as a more general form of the IMC criterion in 8.11.

Recent work by Hjalmarsson *et al.* (1995) has been directed at proving the convergence of such iterative schemes. Hjalmarsson *et al.* (1998) discuss some applications of iterative identification and control to real time control. These include the application to PID control of temperature in a distillation column and the control of a DC-servo motor with backlash.

In this chapter, we focus on developing a method for estimating the control relevant models for MPC under closed loop conditions. Techniques for control relevant MPC are reviewed in Section 8.2. A new interpretation for control relevant pre-filtering is presented followed by discussion of the merits and demerits of the existing methods. Section 8.3 outlines the FIR formulation of the control relevant identification scheme. The control relevant FIR estimate is shown to be independent of the output weighting matrices in MPC. Furthermore, closed loop issues such as constraints, the role of pre-filtering and the impact of nonlinearity are assessed in the remaining part of this Section. Section 8.4 gives the concluding remarks.

8.2 Control Relevant Identification for MPC

Shook *et al.* (1992) considered the following open-loop identification objective

$$J_{IRPI} = \sum_{k=1}^{N-p} \sum_{i=1}^{p} (y(k+i/k) - \hat{y}(k+i))^2 \tag{8.12}$$

which is compatible with the MPC objective for the case $\Gamma = I, \Lambda = 0$. The process was assumed to be of the autoregressive integrated moving average with exogneous input

(ARIMAX) form,

$$\hat{A}(q^{-1})y(k) = \hat{B}(q^{-1})u(k-1) + \frac{C(q^{-1})}{\Delta}a(k).$$

The j-step ahead predictor is derived using the Diophantine identity,

$$\hat{y}(k+j/k) = \frac{F_j}{C}y(k) + \frac{E_j\hat{B}}{C}\Delta u(k+j-1).$$

Substituting the predictor into J_{LRPI} we get

$$J_{LRPI} = \sum_{k=1}^{N-p}\sum_{i=1}^{p}\left[\frac{E_j\hat{A}\Delta}{C}\left(\frac{B}{A} - \frac{\hat{B}}{\hat{A}}\right)u(k+j-1)\right]^2. \tag{8.13}$$

Using Parseval's theorem, a control relevant pre-filter, $L_s(q^{-1})$ is derived such that

$$|L_s(e^{iw})|^2 = \sum_{j=1}^{p}|E_j(e^{iw})|^2.$$

The existence of an unique L_s is guaranteed and spectral factorization is used to obtain the pre-filter. This pre-filter reduces the least squares objective function to the control relevant case,

$$\begin{aligned} J_{LRPI}(y,u) &= J_{LS}(y_f,u_f) \\ y_f(k) &= L_s(q^{-1})y(k), u_f(k) = L_s(q^{-1})u(k). \end{aligned}$$

The least squares objective function with the pre-filtered data is then equivalent to the control relevant identification objective.

Huang and Wang (1999) used the prediction error model to formulate a multi-step ahead prediction error criterion under closed loop conditions. The process is assumed to be of the form

$$y(k) = G(q^{-1})u(k) + H(q^{-1})a(k) \tag{8.14}$$

and a model of the form

$$\hat{y}(k) = \hat{G}(q^{-1})u(k) + \hat{H}(q^{-1})a(k).$$

The optimal p-step ahead predictor based on the PEM model is given by (Ljung, 1987):

$$\begin{aligned} \hat{y}(k/k-p) &= \hat{W}_p\hat{G}u(k) + (1-\hat{W}_p)y(k) \qquad (8.15) \\ \hat{W}_p &= \hat{F}_p\hat{H}^{-1} \\ \hat{F}_p &= \sum_{i=0}^{p-1} g(i)q^{-i} \end{aligned}$$

where $g(i)$ are the impulse response coefficients of $\hat{G}$. The control relevant objective function for identification is the same as in equation 8.12. Using a similar approach (Parseval's identity) a control relevant pre-filter is derived as

$$L_b = \sum_{j=1}^{p} |F_j(e^{iw})|^2. \tag{8.16}$$

The approach of Huang and Wang is more general as it is based on the PEM structure rather than the ARX structure used by Shook *et al.* Huang and Wang show that their approach gives superior MPC performance under closed loop conditions.

An Interpretation of the control relevant pre-filter.

In the following discussion we lend a new interpretation to the control relevant pre-filter. At each sampling instant we consider the p-step ahead prediction errors weighted suitably. Let us assume, $\Gamma = diag(\gamma_1, \gamma_2, ...\gamma_p)$, in general Γ need not be a diagonal matrix,

$$\begin{aligned}
J_{CR} &= \sum_{k=1}^{N-p}\sum_{i=1}^{p}\gamma_i(y(k+i/k) - \hat{y}(k+i))^2 \qquad (8.17)\\
&= \sum_{i=1}^{p}\gamma_i \sum_{k=1}^{N-p}(y(k+i/k) - \hat{y}(k+i))^2 \\
&= \sum_{i=1}^{P}\gamma_i q^{-2(P-i)} \sum_{k=1}^{N-p}(y(k+p/k) - \hat{y}(k+p))^2 \\
&= L^2(q^{-1}) \sum_{k=p+1}^{N}(y(k) - \hat{y}(k))^2 \\
J_{CR} &= L^2(q^{-1}) J_{LS}.
\end{aligned}$$

There exists no real, linear $L(q^{-1})$ such that $L^2(q^{-1}) = \sum_{i=1}^{P}\gamma_i q^{-2(p-i)}$. Let us see how the frequency response of $L^2(q^{-1})$ looks for $p = 5$, $\gamma_i = i/p$

$$L^2(q^{-1}) = (0.2q^{-8} + 0.4q^{-6} + 0.6q^{-4} + 0.8q^{-2} + 1).$$

For very large N, we invoke Parseval's relation to express the LS objective function in the frequency domain,

$$\begin{aligned}
\sum_{k=P}^{N}(y(k) - \hat{y}(k))^2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_{ee}^2(e^{iw})dw \\
J_{CR}(q^{-1}) &= L^2(q^{-1})\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_{ee}^2(e^{iw})dw \\
J_{CR}(e^{iw}) &= \angle L^2(e^{iw})\frac{1}{2\pi}\int_{-\pi}^{\pi}|L^2(e^{iw})|\phi_{ee}^2(e^{iw})dw \\
|J_{CR}(e^{iw})| &= \frac{1}{2\pi}\int_{-\pi}^{\pi}|L^2(e^{iw})|\phi_{ee}^2(e^{iw})dw.
\end{aligned}$$

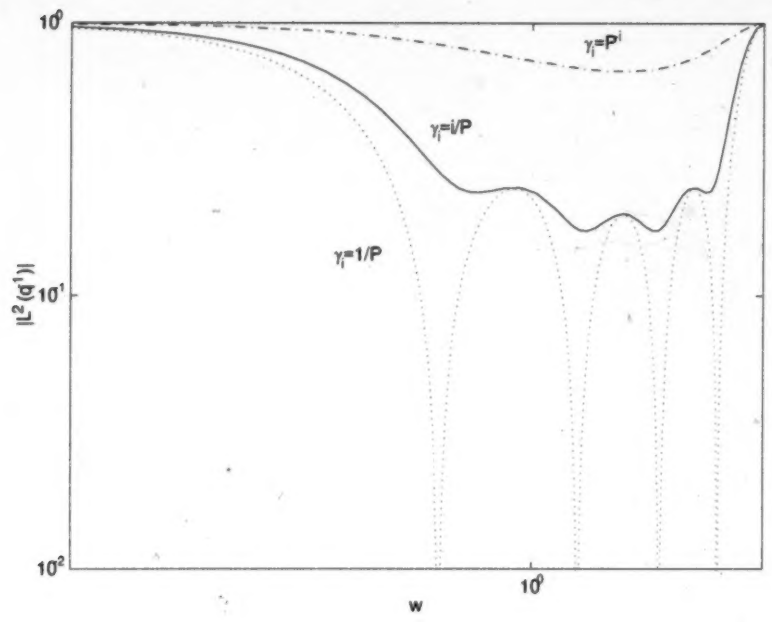


Figure 8.1: *Magnitude response of* $L^2(q^{-1})$ *for different weightings*

Thus the effect of the weighting matrix and the prediction horizon on the control relevant pre-filter is illustrated in an intuitive manner through the preceding derivation.

Banerjee (1996) presents a comparison of different control relevant identification schemes for predictive control. The identification schemes of Shook *et al.* (1992), Rivera *et al.* (1992) and Zang *et al.* (1995) were evaluated in an iterative framework for MPC under closed loop conditions. The third order process of Section 5.4 served as the plant while a first order ARX structure was chosen for the model. The control relevant pre-filter of Shook *et al.* was shown to emphasize the low frequency errors. The control-relevant pre-filter of Rivera *et al.* gave the best achieved performance among the three schemes compared.

Some remarks are in order here with regards to the schemes of Shook et al and Huang and Wang:

- The control relevant identification objective matches the performance degradation term for (i) a particular choice of weighting matrices $\Gamma = I, \Lambda = 0$ in the closed loop case or (ii) the open loop case with $\Gamma = I$. The performance degradation term under closed loop conditions for a general choice of weighting matrices is given by

$$||J(P,C) - J(\hat{P},C)|| = ||y_k - \hat{y}_k||^2_\Gamma + ||\Delta u_k - \Delta \hat{u}_k||^2_\Lambda. \tag{8.18}$$

- It is not clear how the receding horizon nature of MPC will relate to control relevant identification scheme. The receding horizon estimator of Muske *et al.*(1995) does capture this control relevant feature of MPC for estimating the states. The performance degradation term in 8.18 becomes

$$||J(P,C) - J(\hat{P},C)|| = ||y_k - \hat{y}_k||_\Gamma^2 + ||\Delta u(k) - \Delta\hat{u}(k)||_\Lambda^2$$

i.e. we just look at the mismatch between the predicted and the actual input over one time step. In the absence of any input noise or valve type nonlinearity these quantities will be the same. Consequently only the output part of the performance degradation, $||y_k - \hat{y}_k||_\Gamma^2$, remains as the control relevant mismatch.

- The effect of constraints under closed loop conditions is to give a time-varying nature to the control law. How does this affect the implementation of control relevant identification under feedback? Under the assumption of persistent excitation (Ljung, 1987) it is possible to apply direct open-loop identification schemes even under closed loop conditions. This separates MPC from the conventional control schemes where effects of feedback have to accounted for before applying any identification scheme.

- The linearity of a chemical process is more an exception than rule. The derivation of a control relevant pre-filter is of limited use in such cases since it is based on the assumption that the true process is indeed linear and has a structure similar to the model form.

8.3 Control Relevant Identification in the FIR form

In this Section, the control relevant formulation of the finite impulse response is described. An appropriate least squares solution gives the control relevant estimate for the FIR form. The control relevant estimate is shown to be independent of the output weighting. The effect of constraints, nonlinearities and pre-filtering is analyzed in the context closed loop control relevant identification.

8.3.1 Open loop case

In the FIR form the p-step ahead predictor can be expressed as follows,

$$\begin{aligned}
\hat{y}(k+1/k) &= h(1)u(k) + h(2)u(k-1) + ..h(N)u(k-N+1) = \underline{u}_1^T H \\
\hat{y}(k+2/k) &= h(1)u(k+1) + h(2)u(k) + ..h(N)u(k-N+2) = \underline{u}_2^T H \\
&\vdots \\
\hat{y}(k+m/k) &= h(1)u(k+m-1) + h(2)u(k+m-2) + ..h(N)u(k-N+m) = \underline{u}_m^T H \\
&\vdots \\
\hat{y}(k+p/k) &= h(1)u(k+m-1) + h(2)u(k+m-1) + ... + h(p-m)u(k+m-1) \\
&\quad + h(p-m+1)u(k-m-2) + ... + h(N)u(k+P-N) \\
&= \underline{u}_p^T H
\end{aligned}$$

In vector form we have,

$$\begin{aligned}
\hat{Y}_k &= U_k H \\
U_k &= \begin{bmatrix} \underline{u}_1^T \\ \vdots \\ \underline{u}_p^T \end{bmatrix} =
\end{aligned}$$

$$\begin{bmatrix}
u(k) & u(k-1) & & u(k-N+1) \\
u(k+1) & u(k) & & u(k-N+2) \\
\vdots & \ddots & & \\
u(k+m-1) & u(k+m-2) \ldots & & u(k-N+m). \\
u(k+m-1) & u(k+m-1) \ldots & & u(k-N+m+1) \\
\vdots & & \ddots & \\
u(k+m-1) & u(k+m-1) \ldots & u(k+m-1) \quad u(k+m-2) & u(k+P-N)
\end{bmatrix}.$$

Thus the control relevant identification problem in equation 8.17 reduces to identifying the Markov parameters of the system by minimizing the following objective function,

$$\begin{aligned}
&\min_H E[(Y_k - \hat{Y}_k)^T \Gamma (Y_k - \hat{Y}_k)] \\
&= \min_H E[(Y_k - U_k H)^T \Gamma (Y_k - U_k H) \\
&= \min_H \{E(Y_k^T \Gamma Y_k) + E(H^T U_k^T \Gamma U_k H) - 2E(H^T U_k^T \Gamma Y_k)\} \\
&= \min_H \{E(H^T U_k^T \Gamma U_k H) - 2E(H^T U_k^T \Gamma Y_k)\}.
\end{aligned}$$

Assuming H is a deterministic variable, we have the least squares solution as,

$$\begin{aligned}
H_{CR} &= \Sigma_u^{-1} \Sigma_{yu} \qquad (8.19) \\
\Sigma_u &= E(U_k^T \Gamma U_k), \Sigma_{yu} = E(U_k^T \Gamma Y_k).
\end{aligned}$$

H is the impulse response representation of the process. From this one can always back out the state space representation of the process. Note that this is exactly the dual of the control problem without the input term. An interesting property of the control relevant solution is that it is invariant to the output weighting matrix. This is certainly a counter-intuitive result. In order to prove this property we first state a result from linear algebra for generalized inverses of non-square matrices from Campbell and Meyer (1979),

Theorem 8.1. *Suppose* $A \in C^{m\times n}, B \in C^{n\times p}$, *and* $rank(A) = rank(B) = n$. *Then* $(AB)^\dagger = B^\dagger A^\dagger$ *where* $A^\dagger = A^*(AA^*)^{-1}, B^\dagger = (B^*B)^{-1}B^*$, * *denotes complex conjugate transpose.*

Therefore we have

$$\begin{aligned}
\Sigma_u &= E(U_k^T \Gamma U_k) = E(U_k^T \Gamma^{1/2}\Gamma^{1/2} U_k) = E(\tilde{U}_k^T \tilde{U}_k) = \tilde{U}^T\tilde{U} \\
\Sigma_{yu} &= E(U_k^T \Gamma Y_k) = E(\tilde{U}_k^T \tilde{Y}_k) = \tilde{U}^T\tilde{Y} \\
\tilde{U} &= [\Gamma^{1/2}U_1\ \Gamma^{1/2}U_2\ ...\ \Gamma^{1/2}U_N] = \tilde{\Gamma}^{1/2}[U_1\ U_2\ ...\ U_N] = \tilde{\Gamma}^{1/2}U \\
\tilde{Y} &= \tilde{\Gamma}^{1/2}[Y_1\ Y_2\ ...\ Y_N] = \tilde{\Gamma}^{1/2}Y \\
\tilde{\Gamma}^{1/2} &= diag(\Gamma^{1/2}...\Gamma^{1/2}) \\
&\therefore\ \Sigma_u^{-1}\Sigma_{yu} = \tilde{U}^\dagger\tilde{Y}
\end{aligned}$$

Using the earlier result on pseudo-inverses, we have assuming full row rank for U.

$$\tilde{U}^\dagger = (\tilde{\Gamma}^{1/2}U)^\dagger = U^\dagger(\tilde{\Gamma}^{1/2})^\dagger = U^\dagger\tilde{\Gamma}^{-1/2}$$

since $\tilde{\Gamma}^{1/2}$ is a square matrix with full rank. Therefore we have,

$$\begin{aligned}
\Sigma_u^{-1}\Sigma_{yu} &= U^\dagger\tilde{\Gamma}^{-1/2}\tilde{\Gamma}^{1/2}Y \\
&= U^\dagger Y.
\end{aligned}$$

Thus the control relevant FIR estimate is independent of the output weighting matrix. Let us compare this solution to the conventional least squares estimation of the impulse response model of the system,

$$\begin{aligned}
\hat{y}(k+1) &= h(1)u(k) + h(2)u(k-1) + ..h(N)u(k-N+1) = \mathfrak{u}_1^T H \\
\hat{y}(k+2) &= h(1)u(k+1) + h(2)u(k) + ..h(N)u(k-N+2) = \mathfrak{u}_2^T H \\
&\vdots \\
\hat{y}(k+N_1) &= h(1)u(k+N_1-1) + h(2)u(k+N_1-2) + ..h(N)u(k+N_1-N) = \mathfrak{u}_{N_1}^T H
\end{aligned}$$

i.e.

$$\hat{Y} = UH.$$

Therefore the conventional least squares problem is

$$\begin{aligned} & \min_H (Y - \hat{Y})^T (Y - \hat{Y}) \\ = \; & \min_H (Y - UH)^T (Y - UH). \end{aligned}$$

The least square solution is given by

$$U_{LS} = (U^T U)^{-1} U^T Y. \tag{8.20}$$

What is the difference between the conventional least squares solution and the control relevant estimate of the Markov parameters? The answer, as the subsequent example reveals, is very little. The reason is that for a linear, stable process there is no under parametrization with the FIR form. As a result there are no bias errors. Hence the question of distributing the bias errors in a control relevant fashion never really arises. As far as the variance of the respective estimates is concerned it can be shown that the control relevant estimates show lesser variation when compared with the least squares solution. Figure 8.2 shows the comparison of the control relevant estimates with the least squares estimate for the third order process in Chapter 5. They are almost identical. The control relevant estimate was found to have a covariance matrix which was an order of magnitude lower than the least squares estimate.

8.3.2 MIMO extensions

The FIR approach extends very naturally to the MIMO case. We illustrate the MIMO problem for a 2×2 case. The generalization to higher dimensions is straightforward,

$$\begin{aligned} y_1(k) &= \sum_{i=1}^{N} h_{11}(i) u_1(k-i) + \sum_{i=1}^{N} h_{12}(i) u_2(k-i) \\ &= H_{11}^T \mathfrak{u}_1(k) + H_{12}^T \mathfrak{u}_2(k) \\ y_2(k) &= H_{21}^T \mathfrak{u}_1(k) + H_{22}^T \mathfrak{u}_2(k) \\ y(k) &= \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} H_{11}^T & H_{12}^T \\ H_{21}^T & H_{22}^T \end{bmatrix} \begin{bmatrix} \mathfrak{u}_1(k) \\ \mathfrak{u}_2(k) \end{bmatrix} \\ y(k)^T &= \begin{bmatrix} \mathfrak{u}_1^T(k) & \mathfrak{u}_2^T(k) \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}. \end{aligned}$$

The conventional least squares problem is equivalent to finding the solution to

$$\begin{bmatrix} y(1)^T \\ y(2)^T \\ \\ y(N_1)^T \end{bmatrix} = \begin{bmatrix} \mathfrak{u}_1^T(1) & \mathfrak{u}_2^T(1) \\ \mathfrak{u}_1^T(2) & \mathfrak{u}_2^T(2) \\ \\ \mathfrak{u}_1^T(N_1) & \mathfrak{u}_2^T(N_1) \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

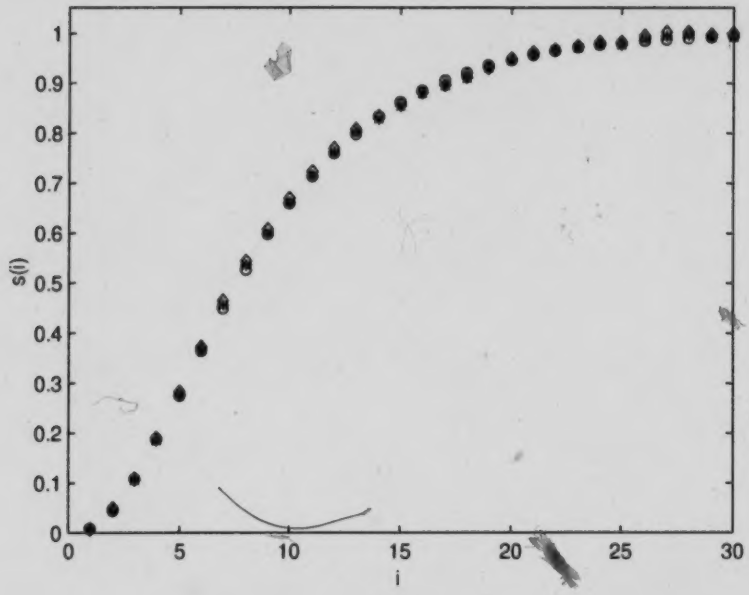


Figure 8.2: *Comparison of control relevant FIR identification with conventional least squares method. True impule response (o), control relevant model (*) and conventional least squares estimate* ($\triangle$)

For the control relevant case we have the prediction vector $\hat{Y}_k$ instead of $y(1)$ for each of the outputs and the matrices $\hat{U}_k$ for each input in place of $\mathrm{u}_1^T(1)$ *etc.*

8.3.3 Closed loop issues

We presented an industrial application of correlation analysis to closed loop identification in chapter 7. Here we will try to analyze the utility of using the FIR model in closed loop identification with the pre-whitening filter. Consider the closed loop system given in figure 8.3. The closed loop relations are given by,

$$y = Tr + SPw + Sv$$

where w is the excitation at the input, we assume an excitation sequence of the form,

$$w = Fe$$

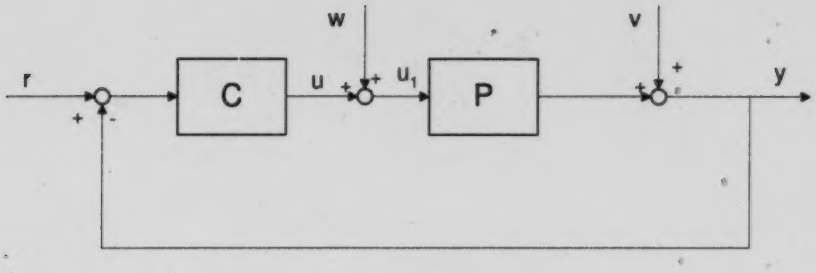


Figure 8.3: *The feedback structure*

where e is a white noise process. Take the case where $r = 0, v = 0$

$$\begin{aligned} u_1 &= u + w \\ y &= Pu_1 \\ u_1 &= Sw \\ &= SFe. \end{aligned}$$

If a pre-whitening filter is estimated for u_1, as $Lu_1 = \hat{e}$, then the question is under what conditions is $L^{-1} = SF$? If this holds than the pre-whitening filter accounts for the feedback and should still be able to give an estimate of the open-loop relationship between the inputs and outputs. Figure 8.4 shows the magnitude response of the pre-whitening filter for a closed loop system comprising of the third order process (equation 5.16) and a proportional feedback controller, $C = 6$. A white-noise type dither signal ($F = 1$) was used to excite the closed loop. In this case L^{-1} matches the sensitivity function response very closely. Figure 8.5 shows the role of the pre-whitening filter on closed loop identification of the FIR model. When the pre-whitening filter is used the FIR model between the filtered input and output matches the true step response. Without the pre-whitening filter the FIR model shows influence of feedback and estimates a negative gain with a highly oscillatory response.We also investigated the role of the pre-whitening filter on estimation of output (OE) models which are essentially the same as FIR models in terms of structure. Figures 8.6 and 8.7 show the role of the pre-whitening filter on the OE structure in the time and frequency domains respectively. Once again without the pre-whitening filters the OE model shows bias. The small bias in the OE model, with pre-whitening, at steady state could be attributed to the fact that the sensitivity function acts as a high pass filter and the pre-whitening filter is not able to reconstruct the low frequency signals with desired accuracy.

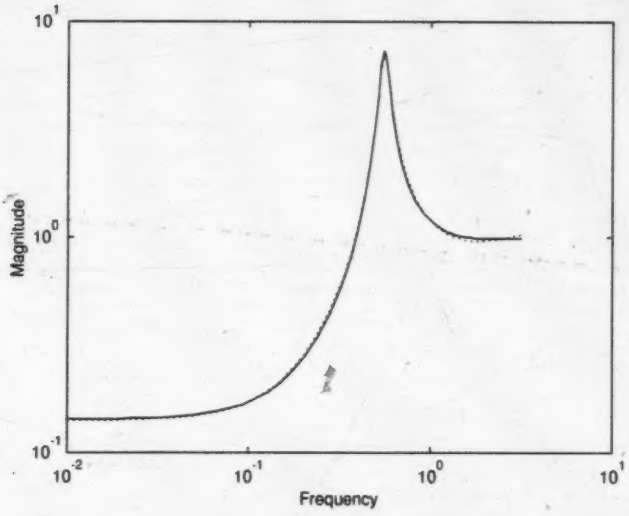


Figure 8.4: *The magnitude response of the pre-whitening filter (dashed line) and the sensitivity function (solid line).*

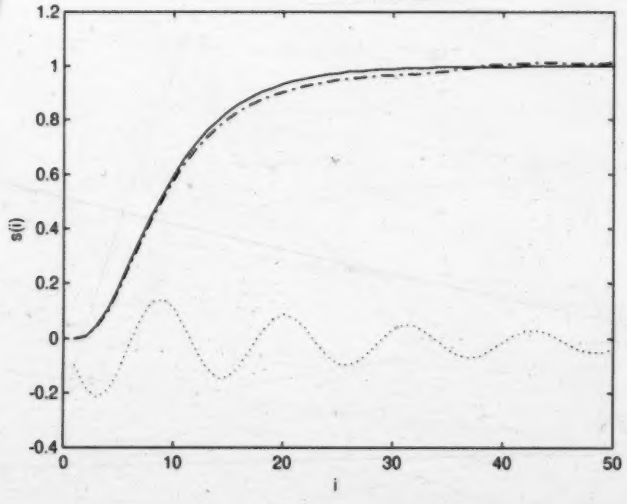


Figure 8.5: *Estimation of the step response coefficients using correlation analysis under closed loop conditions with pre-whitening (dash-dotted) and without pre-whitening (dotted) shown with the true step response (solid).*

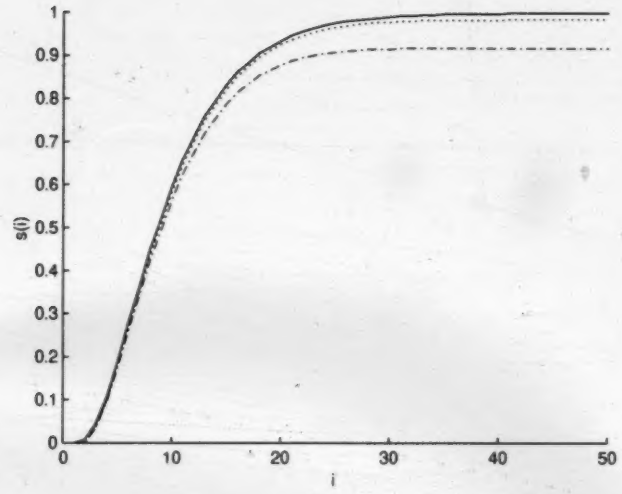


Figure 8.6: *The effect of pre-whitening on the OE structure: OE with pre-whitening (dotted line), without pre-whitening (dashdotted line) and the true response (solid line).*

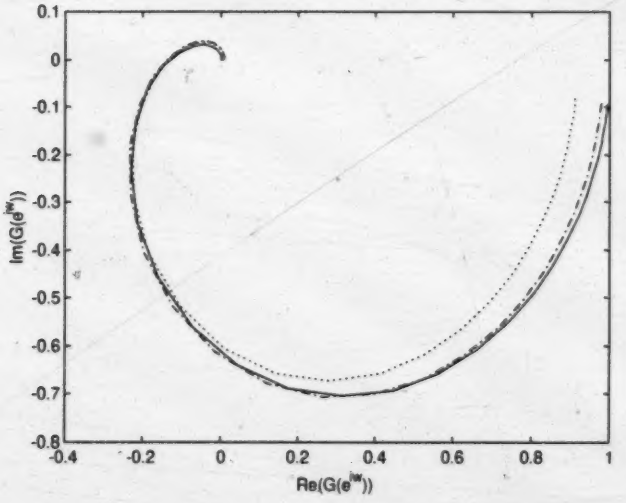


Figure 8.7: *The frequency domain fits for the OE models with and without pre-whitening. OE with pre-whitening (dotted line), without pre-whitening (dashdotted line) and the true response (solid line).*

Effect of constraints

The presence of constraints in a controller changes the feedback system from being a linear time invariant one to a linear time varying one. In effect it modifies the linear correlation structure under feedback. One cannot express the closed loop relationships in terms of transfer functions. From an identification standpoint this is a desirable property since it enables application of direct input-output identification under feedback. The two conditions necessary for applicability of direct input-output identification in the closed loop are:

1. The controller is time-varying or changing.

2. The presence of a persistent excitation signal.

Condition 1 is satisfied by a constrained controller when the active set of constraints is changing at every sampling instant. This is a very useful fact for estimating models in closed loop. The constraints can act as an extra degree of freedom for the design of closed loop experiments. For example, for the duration of the identification run, the constraints can be altered in such a way that the excitation sequence hits a different constraint every sampling instant. The following example illustrates the role of constraints in closed loop identification.

In the closed loop system of the previous example, we added rate and amplitude constraints on the proportional controller:

$$-2 \leq u(k) \leq 2, -0.5 \leq \Delta u(k) \leq 0.5. \tag{8.21}$$

Figure 8.8 shows the effect of constraints on direct identification without any pre-filtering. For the unconstrained case there is non-trivial bias whereas for the constrained case there is very little mismatch. We also studied the effect of time span for which constraints are active on the direct identifiability scheme. In figure 8.9 the dashdotted line refers to the constraints in equation 8.21. The constraints are active for over 85% of the time, during the closed loop experiment. Subsequently the rate constraints were relaxed to $-1 \leq \Delta u(k) \leq 1$ and the constraints were active for 66% of the run. The direct identification scheme is less accurate in this case as shown by the dotted line in figure 8.9. When the constraints were further relaxed to $-5 \leq \Delta u(k) \leq 5$ the active period reduces to 15% (see the dashed line in figure 8.9) and the result from the direct input-output identification scheme is close to the unconstrained case.

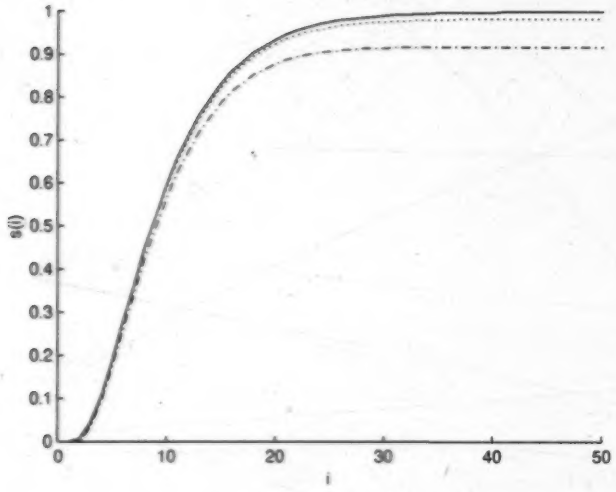


Figure 8.8: *Effect of constraints on closed loop identification: direct identification with constraints (dashed line), without constraints (dashdotted line). The solid line gives the true step response.*

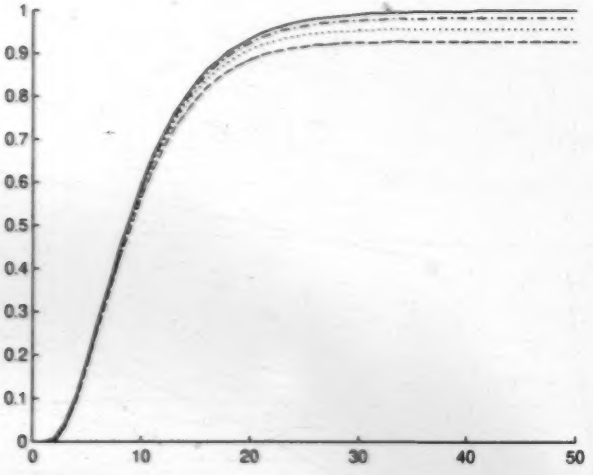


Figure 8.9: *Effect of constraint activation on closed loop identification.*

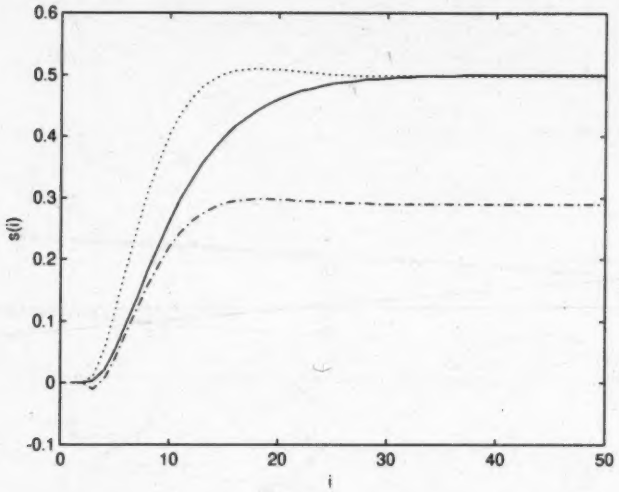


Figure 8.10: *Effect of valve nonlinearity on closed loop identification schemes.*

Impact of nonlinearity

As with constraints, the presence of nonlinearity also makes the linear time invariant closed loop relationships infeasible. We studied the effect of a simple backlash type nonlinearity on the direct identification scheme with and without pre-filtering. Figure 8.10 shows the effect of backlash in the control valve. The direct identification without any filtering (dashed line) is able to capture the steady state part of the process response accurately. In the presence of filtering with a pre-whitening filter (dashdotted line) the model shows a large bias at steady state but gives a better fit at higher frequencies or initial times.

8.4 Conclusions

The main contributions of this chapter can be summarized as follows:

- A control relevant identification scheme is formulated for the finite impulse response model. It was shown that the output weightings have no effect on the outcome of the control relevant estimation scheme. The conventional least squares method and proposed scheme were shown to give similar results on a simulation problem. The proposed scheme led to estimates with lesser variance.
- The effect of the pre-whitening filter in closed loop identification with the FIR and

OE forms was analyzed. It was shown that the pre-whitening filter is equivalent to the sensitivity function for the case of high signal to noise ratio and persistent excitation.

- The role of constraints and nonlinearity on direct closed loop identification was demonstrated via an illustrative example. It was shown that activation of constraints enables application of direct identification under closed loop conditions. The frequency with which the active constraint set changes is important to the accuracy of identification in closed loop. The impact of nonlinearity is not so obvious. A backlash type of nonlinearity is used to illustrate the dangers of using filtering under feedback.

We have taken some steps towards implementing iterative control and identification under closed loop conditions. Further work is required to design a robustly performing iterative MPC scheme based on the FIR form. In theory, the iterative scheme provides an ideal solution to the continuous maintenance and improvement of model predictive control applications.

Chapter 9

Conclusions and Recommendations

9.1 Concluding remarks

The main contributions of this thesis are:

- A tutorial overview of model predictive control and the prevalent research directions. The basic principles of MPC are explained in a transparent manner using illustrative examples. Current research in the design of robust MPC is presented.

- Application of linear matrix inequalities in the design and tuning of model predictive controllers. The LMI approach, as shown by Kothare *et al.* (1996), gives an elegant framework for casting and solving the robust MPC problem efficiently. We present some real time applications of the LMI based MPC and give a new way of choosing the MPC weighting matrices via the LMI framework.

- Design of linear and nonlinear MPC in the partial least squares framework. The PLS framework is useful in modeling correlated data. The dynamic extensions of PLS based models are shown to have some theoretically desirable properties. Linear and nonlinear MPC using PLS models is proposed, and demonstrated via simulation and real time applications.

- Development of a new assessment technique for MPC based on comparison of the design and achieved objective functions. Using the state space form it was shown how to estimate the design objective function analytically. Receding horizon was shown to lead to a loss in performance.

- Quantifying the effect of uncertainty, constraints, mismatch in the disturbance model, nonlinearity on loss in controller performance. This presents an important step towards diagnosing poor performance. Closed loop tests based on external excitation

can help eliminate some of the above factors. Future work needs to be directed developing factor specific tests that will aid the diagnosis procedure.

- Application of the performance assessment and diagnostic tools to two industrial DMC applications. The DMC structure was analyzed and effect of the LP part on closed loop stability was established. Several practical measures and diagnostics were emphasized in this study. The role of PID tuning and constraint activation in determining DMC performance was illustrated.

- A control relevant identification method for the FIR form was proposed and evaluated. The role of filtering under closed loop conditions was ascertained. The effect of constraints and nonlinearity on model estimation, under feedback, was highlighted.

The proposed synthesis and analysis tools were evaluated on a variety of real time and simulation examples:

- The LMI approach was implemented on two pilot scale process - a light bulb temperature control process (SISO) and a stirred tank heater process (MIMO).

- The linear PLS based MPC was also evaluated on the stirred tank heater. The nonlinear PLS based MPC was applied on a simulated pH-level control process.

- The proposed assessment and diagnostic tools were implemented on (i) a commercial QDMC application at the Scotford refinery of Shell Canada and (ii) two DMC applications at the Mizushima plant of Mitsubishi Chemicals, Japan.

- Apart from these applications, appropriate simulation examples were used to illustrate the design and analysis tools in a tutorial manner.

Thus this thesis has taken some steps towards improving the design and analysis of model predictive controllers. The presented techniques are simple and appealing to the industrial practitioner. The experience gathered in analyzing industrial applications will lead to better understanding of the industrial technology for MPC.

9.2 Future research directions

A wealth of tuning and design methods are available for MPC. However, relatively little work has been done in the area of analyzing performance of existing MPC applications. Some progress was made in this work towards establishing performance metrics suited for

MPC. A great deal of work needs to be done in the area of diagnostics, in establishing the contributing factors to performance degradation and the extent of the contribution. A MPC application needing little maintenance would be one which monitors its own performance and can iterate between the design and analysis phase intelligently.

A self learning controller which finds out when things are not right, establishes the underlying causes and implements corrective measures or at least makes suggestions would be the dream of any industrial practitioner. The iterative control and identification (Van Den Hof and Schrama, 1995) framework is ideal for such a controller. The identification scheme where the controller itself injects the dither signal automatically (Genceli and Nikolaou, 1996) is also well suited for purposes of re-identifying the process or merely establishing the degrading factors such as disturbances, constraints, etc.

Towards this end, some of the problems that need to be investigated are:

- Establishing systematic methodologies for diagnosis of poor controller performance. This includes answering questions such as: when are disturbances responsible for a loss in performance? Is it a combination of disturbances and model plant mismatch?
- Development of factor-specific closed loop tests. For example, injecting an excitation signal at the input may help in determining the model plant mismatch. Other kinds of signals may aid detection of nonlinearities.
- Studying the effect of nonlinearity on performance of linear mode-based controllers. In other words, what is the loss in performance due to linear a controller working on a nonlinear process?
- The state space framework provides a powerful tool for designing and analyzing MPC. Development of control relevant state space estimation techniques would prove useful in implementing iterative MPC schemes based on the state space model.
- Techniques for on-line re-tuning of model predictive controllers. If the performance is not up to the mark, what tuning parameters need to be changed and by how much? This is not a trivial problem and often addressed in a heuristic manner in industrial practice.

Bibliography

Allgower, F. (1995). Definition and computation of a nonlinearity. In: *IFAC Symposium on Nonlinear Control System Design*. Tahoe City, CA.

Astrom, K. J. and B. Wittenmark (1991). *Computer-Controlled Systems: Theory and Design*. Prentice Hall. NY.

Badmus, O. O., S. L. Shah and Fisher D. G. (1996). Case studies in system identification. Technical report. Dept. of Chem. Eng., University of Alberta.

Bannerjee, P. (1996). Robustness Issues in Long Range Predictive Control. PhD thesis. Dept. of Chemical Eng., University of Alberta, Edmonton.

Bannerjee, P. and S. L. Shah (1995). The role of signal processing methods in the robust design of predictive control. *Automatica* **31**, 681–695.

Bemporad, A. and Morari (1998). Control of systems integrating logic, dynamics and constraints. Presented at the AICHE Annual Meeting, Miami.

Bhat, N. and T. J. McAvoy (1990). Use of neural nets for dynamic modeling and control of chemical process systems. *Comput. Chem. Eng.* **14**, 573–583.

Biegler, L. T. (1984). Solution of dynamic optimization problems by successive quadratic programming an orthogonal collocation. *Comput. Chem. Eng.* **8**, 243–248.

Bitmead, R. R., M. Gevers and M. Wertz (1990). *Adaptive Optimal Control: The Thinking Man's GPC*. Prentice Hall.

Boyd, S., L. Vandenberghe and M. Grant (1994). Efficient convex optimization for engineering design. In: *Proceedings of IFAC Symp. on Robust Control Design*. Rio De Janerio, Brazil.

Boyd, S. P. and C. H. Barrat (1991). *Linear Controller Design: Limits of Performance*. Prentice Hall. NY.

Brengel, D. D. and W. D. Seider (1989). Multistep nonlinear model predictive controller. *Ind. Eng. Chem. Res.* **28**, 1812–1822.

Campbell, S. L. and C. D. Meyer Jr. (1979). *Generalized Inverses of Linear Transformations*. Dover. New York.

Campo, P.J. and M. Morari (1987). Robust model predictive control. In: *Proc. American Control Conf.*. Minneapolis, MN. pp. 1021–1026.

Chen, T. and B. A. Francis (1995). *Optimal Sampled-Data Control Systems*. Springer Verlag. NY.

Clarke, D. W. and C. Mohtadi (1989). Properties of genralized predictive control. *Automatica* **25**, 859–875.

Clarke, D. W., C. Mohtadi and P. S. Tuffs (1987). Generalized predictive control - part 1 and 2. *Automatica* **23**, 137–160.

Cutler, C. R. and B. L. Ramaker (1990). Dynamic matrix control - a computer control algorithm. In: *Proceedings of Joint American Control Conference*. San Fransisco, USA.

Dahleh, M. and I.J. Diaz-Bobillo (1995). *Control of Uncertain Systems: A Linear Programming Approach*. Prentice Hall. Englewood Cliffs, NJ.

Dahlin, E. B. (1967). Designing and tuning of digital controllers. *Instr. and Control Sys.* **41**, 77.

de Oliveira, S. L., V. Nevistic and M. Morari (1995). Control for nonlinear systems subject to input constraints. In: *Proceedings of the IFAC NOLCOS Symposium*. Lake Tahoe, CA.

Desborough, L. and T. Harris (1992). Performance assessment measure for univariate feedback control. *Can. J. Chem. Eng.* **70**, 605–616.

Doyle III, F. J., B. A. Ogunnaike and R. K. Pearson (1995). Nonlinear model-based control using second order Volterra models. *Automatica* **31**, 697–714.

Doyle, J. C., B. A. Francis and A. R. Tannenbaum (1992). *Feedback Control Theory*. MacMillan Publishing Co.. NY.

Economou, C. G. and M. Morari (1986). Internal model control 5: Extension to nonlinear systems. *Ind. Eng. Chem. Res.* **25**, 403–411.

Emoto, G.and Ota, Y., H. Matsuo, M. Ogawa and I. B. Tjoa (1996). On the integration of real-time optimization and control for ethylene plant. AIChE Annual Meeting, Chicago.

Gahinet, P., A. Nemirovsky, A. J. Laub and M. Chilai (1995). *LMI Control Toolbox: For use with MATLAB*. MathWorks Inc.

Garcia, C. E. (1984). Quadratic dynamic matrix control of nonlinear processes: An application to a batch reaction process. AIChE Annual Meeting, San Fransisco, CA.

Garcia, C. E. and M. Morari (1982). Internal model control. 1. a unifying review and some new results. *I&EC Proc. Des. and Dev.* **24**, 308–323.

Garcia, C. E., D. M. Prett and M. Morari (1989). Model predictive control: Theory and practice - a survey. *Automatica* **25**, 335–348.

Gattu, G. and E. Zafirou (1992). Nonlinear quadratic dynamic matrix control with state estimation. *Ind. Eng. Chem. Res.* **31**, 1096–1104.

Genceli, H. and M. Nikaloau (1996). A new approach to constrained model predictive control with simultaneous identication. *AIChEJ.* **42**, 2857–2868.

Genceli, H. and M. Nikolaou (1993). Robust stability analysis of l_1-norm model predictive control. *AIChE J.* **39**, 1954–1965.

Gevers, M. (1993). Towards a joint design of identification and control?. In: *Essays on Control: Perspectives in the Theory and its Applications*. pp. 111–151. Birkhauser. Boston.

Goodwin, G. and K. Sin (1984). *Adaptive Filtering, Prediction and Control*. Prentice-Hall. NY.

Guay, M., P. J. McLellan and D. W. Bacon (1995). Measurement of nonlinearity in chemical process control systems: The steady state map. *Can. J. Chem. Eng.* **73**, 868–882.

Gustafsson, F. and S. F. Graebes (1998). Closed-loop performance monitoring in the presence of system changes and disturbances. *Automatica* **34**, 1311–1326.

Haber, R. and H. Unbehauen (1990). Structure identification of nonlinear dynamic systems - a survey on input/output approaches. *Automatica* **26**, 651–677.

Harris, T. J. (1989). Assessment of closed loop performance. *Can. J. Chem. Eng.* **67**, 856–861.

Harris, T. J., F. Bourdeau and J. F. MacGregor (1997). Performance assessment of multivariate feedback controllers. *Automatica* **32**, 1505–1518.

Henson, M. A. and D. E. Seborg (1994). Adaptive nonlinear control of a ph neutralization process. *IEEE Trans. Aut. Control* **18**, 552–564.

Hernandez, E. and Y. Arkun (1993). Control of nonlinear systems using polynomial arma models. *AIChE J.* **39**, 446–460.

Hjalmarsson, H., M. Gevers, S. Gunnarsson and O. Lequin (1998). Iterative feedback tuning: Theory and applications. *IEEE Control Systems* **18**, 26–41.

Hjalmarsson, H., S. Gunnarsson and M. Gevers (1995). Optimality snd sub-optimality of iterative identification and control design schemes. In: *American Control Conf.*. Vol. 4. pp. 2559–2563.

Howie, B. (1995). Application manual on hcu recycle surge drum level control. Technical report. Shell Canada, Scotford Refinery.

Huang, B. (1997). Multivariate Statistical Methods for Control Loop Performance Assessment. PhD thesis. Dept. of Chem. Eng., University of Alberta, Edmonton.

Huang, B. and Z. Wang (1998). The role of data pre-filtering for integrated identification and model predictive control. Accepted for presentation at the IFAC World Congress, Beijing, China.

Huang, B., S. L. Shah and K. Y. Kwok (1997). Good, bad or optimal? performance assessment of mimo processes. *Automatica* **33**, 1175–1183.

Jerome, N. F. and W. H. Ray (1986). High-performance multivariable control strategies for systems having multiple delays. *AIChE J.* **32**, 914.

Kammer, L. C., R. R. Bitmead and P. L. Barlett (1996). Signal-based testing of lq-optimality of controllers. In: *Proceedings of the* 35^{th} *Conference on Decision and Control.* Kobe, Japan. pp. 3620–3624.

Karmarkar, N. (1984). A new polynomial time algorithm for linear programming. *Combinatorica* **4**(4), 373–395.

Kaspar, M. H. and W. H. Ray (1992). Chemometric methods for process monitoring. *AIChEJ.* **38**, 1593–1608.

Kaspar, M. H. and W. H. Ray (1993). Dynamic pls modeling for process control. *Chem. Eng. Sci* **48**, 3447–3461.

Kesavan, P. and J. H. Lee (1997). Diagnostic tools for multivariable model-based control systems. *Ind. Eng. Chem. Res.* **36**, 2725.

Kothare, M. V., V. Balakrishnan and M. Morari (1996). Robust constrained model predictive control using linear matrix inequalities. *Automatica* **32**, 1361–1379.

Kozub, D. and C. E. Garcia (1993). Monitoring and diagnosis of automated controllers in the chemical process industries. Presented at the AIChE Annual Meeting.

Kozub, D. J. (1996). Controller performance monitoring and diagnosis experiences and challenges. In: *Proceedings of CPC-V.* Lake Tahoe, CA.

Kresta, J. (1992). "Applications of Partial Least Squares Regression". PhD thesis. McMaster University.

Kwon, W. and A. Pearson (1977). A modified quadratic cost problem and feedback stabilization of a linear system. *IEEE Trans. Aut. Cont.* **22**, 838–842.

Lakshminarayanan, S., S. L. Shah and K. Nandakumar (1995). Identification of Hammerstein models using multivariate statistical tools. *Chem. Eng. Sci.* **50**, 3599–3613.

Lakshminarayanan, S., S. L. Shah and K. Nandakumar (1997a). Modeling and control of multivariable systems: The dynamic projection to latent structures approach. *AIChEJ.* **43**, 2307.

Lee, J. H. (1996). Recent advances in model predictive control and other related areas. In: *Proceedings of CPC-V.* Lake Tahoe, CA.

Lee, J. H., M. Morari and C. E. Garcia (1994). State-space interpretation of model predictive control. *Automatica* **30**, 707–717.

Lee, J., W. H. Kwon and J. Choi (1998). On stability of constrained receding horizon control with finite terminal weighting matrix. *Automatica* **34**, 1607–1612.

Lee, W. S. B., B. D. O. Anderson, R. L. Kosut and I. M. Y. Mareels (1993). A new approach to adaptive robust control. *Int. J. Adaptive Control and Signal Proc.* **7**, 183–211.

Li, S., K. Y. Lim and D. G. Fisher (1989). A state space formulation for model predictive control. *AIChEJ.* **35**, 241.

Li, W. C. and L. T. Biegler (1988). Process control strategies for constrained nonlinear systems. *Ind. Eng. Chem. Res.* **27**, 1421–1433.

Li, W. C. and L. T. Biegler (1989). Multistep newton type control strategies for constrained nonlinear processes. *Chem. Eng. Res. Des.* **67**, 562.

Ljung, L. (1987). *System Identification: Theory for the User.* Prentice-Hall. Englewood Cliffs, NJ.

Maciejowski, J. M. (1989). *Multivariate Feedback Design.* Addison-Welsley. NY.

Manner, B. R., F. J. Doyle III and R. K. Pearson (1996). Nonlinear model predictive control of a simulated multivariable polymerization reactor. *Automatica* **32**, 1285–1301.

Michalska, H. and D. Q. Mayne (1992). Moving horizon observers. In: *IFAC Symposium on Nonlinear Control System Design.* Bourdreaux, France. pp. 576–681.

Morari, M. and E. Zafiriou (1989). *Robust Process Control.* Prentice-Hall. NY.

Muske, K. R., J. B. Rawlings and J. H. Lee (1993). Receding horizon recursive estimation. In: *American Control Conf.*. San Fransisco, CA. pp. 900–904.

Mutha, R. K. (1990). Constrained long range predictive control. Master's thesis. Dept. of Chem. Eng., University of Alberta.

Nesterov, Y. and A. Nemirovsky (1994). *Interior Point Polynomial Methods in Convex Programming.* Vol. 13 of *Studies in Applied Mathematics.* SIAM, Philadelphia.

Nikolaou, M. (1993). When is nonlinear modeling necessary. In: *American Control Conference.* San Francisco, CA. pp. 910–914.

Nomikos, P. and J.F. MacGregor (1995). "Multivariate SPC Charts for Monitoring Batch Processes". *Technometrics* **37**(1), 41–59.

Norquay, S. J., A. Palazoglu and J. A. Romagnoli (1997). Unconstrained and constrained model predictive control using wiener models: An application to ph neutralizatio. In: *IFAC Symp. on Adv. Control Chem. Proc.*. p. 541.

Norquay, S. J., A. Palazoglu and J. A. Romangoli (1996). Nonlinear model predictive control of pH neutralization using Wiener models. In: *Proceedings of the* 13^{th} *IFAC World Congress, San Fransisco, CA*.

Ogunnaike, B. A. and W. H. Ray (1979). Multivariable controller design for linear systems having multiple delays. *AIChEJ.* **25**, 1043.

Patwardhan, A. A., J. B. Rawlings and T. Edgar (1990). Nonlinear model predictive control. *Chem. Eng. Commun.* **87**, 123–141.

Patwardhan, S. C. and K. P. Madhavan (1993). Nonlinear model predictive control using second order model approximation. *Ind. Eng. Chem. Res.* **32**, 334–344.

Phatak, A., P. M. Reilly and A. Penlidis (1993). An approach to interval estimation in partial least squares regression: A tutorial. *Anal. Chem. Acta.* **277**, 495–501.

Polak, E. and T. H. Yang (1993). Moving horizon control of linear systems with input saturation and plant uncertainty, parts 1 and 2. *Int. J. Control* **58**, 613–663.

Qi, K. (1997). Robust Design of Model Predictive Controllers. PhD thesis. University of Alberta, Edmonton.

Qin, J. S. and T. A. Badgewell (1996). An overview of industrial model predictive control technology. In: *Fifth Int. Conf. on Chem. Proc. Control* (J. C. Kantor and C. E. Garcia, Eds.). Vol. 93. AIChE Symposium Series. p. 232.

Qin, J. S. and T. J. McAvoy (1992). A data-based process modeling approach and its application. In: *Proceedings of the IFAC Conference on Dynamics and Control of Chemical Reactors (DYCORD+'92).*

Ralhan, S. (1998). Robust model predictive control for linear finite impulse response. Master's thesis. Rice University. Houston.

Rawlings, J. B. and K. R. Muske (1993). The stability of constrained receding horizon controllers. *IEEE Trans. Aut. Control* **AC-38**, 1512–1516.

Richalet, J., A. Rault, J. L. Testud and J. Papan (1978). Model predictive heuristic controller applications to industrial processes. *Automatica* **14**, 413–428.

Ricker, N. L. (1988). The use of biased-least squares for parameters in discrete time impulse response models. *Ind. Eng. Chem. Res.* **27**, 343–350.

Rivera, D. E., J. F. Pollard and C. E. Garcia (1992). Control-relevant pre-filtering: A systematic design approach and case-study. *IEEE Trans. Autom. Control* **37**, 964–974.

Rouhani, R. and R. K. Mehra (1982). Model algorithmic control. *Automatica* **18**, 401–414.

Schrama, R. J. P. and O. H. Bosgra (1993). Adaptive performance enhancement by iterative identification and control design. *Int. J. Adaptive Control & Signal Processing* **7**, 475–487.

Schwarm, A. T. and M. Nikolaou (1998). Robust output constraint satisfaction in mpc: Optimizing under uncertainty. Presented at the AIChE Annual Meeting, Miami, Fl.

Scokaert, P. O. M. and J. B. Rawlings (1998). Constrainted linear quadratic regulation. *IEE Trans. Aut. Cont.* **43**, 1163.

Shah, S. L., C. Mohtadi and D. W. Clarke (1987). Multivariable adaptive control without a prior knowledge of the delay matrix. *Syst. & Control Letters* **9**, 295.

Shook, D. S., C. Mohtadi and S. L. Shah (1992). A control-relevant identification strategy for gpc. *IEEE Trans. Aut. Control* **37**, 975–980.

Skogestad, S. and I. Postlethwaite (1996). *Multivariate Feedback Control: Analysis and Design*. John-Wiley & Sons. NY.

Sripada, N. R. and D. G. Fisher (1985). Multivariable optimal constrained control algorithm (mocca): Part 1. formulation and application. In: *Proceedings of International Conference on Modeling and Control*. Hangzhou, China.

Stack, A. J. and F. J. Doyle III (1997). The optimal control structure: An approach to measuring control-relevant nonlinearity. *Comp. Chem. Eng.* **21**, 998–1009.

Stanfelj, N., Marlin T. E. and J. F. MacGregor (1993). Monitoring and diagnosing process control performance: The single-loop case. *Ind. Eng. Chem. Res.*

Tan, K. K., Q. G. Wang, T. H. Lee and Q. Bi (1996). New approach to analysis and design of smith-predictor controllers. *AIChE J.* **42**, 1793.

Thake, A., J. F. Forbes and P. J. McLellan (1997). Approximation of high dimensional multivariable controllers using semi-definite programming. In: *IFAC Symp. Adv. Control Chem. Proc.*. pp. 664–669.

Toker, O., J. Chen and L. Qiu (1997). Limitations on optimal tracking performance of discrete-time systemsms. In: *American Control Conference*. pp. 3887–3991.

Tsiligiannis, C. A. and S. A. Svoronos (1988). Dynamic interactors in multivariable process control - i. the general time delay case. *Chem. Eng. Sci.* **43**, 339.

Tyler, M. L. and M. Morari (1995). Performance monitoring of control systems using likelihood methods. In: *Proc. American Control Conf.*. pp. 1245–1249.

Tyler, M. L. and M. Morari (1996). Performance monitoring of control systems using likelihood methods. *Automatica* **32**, 1145–1162.

Van den Hof, P. M. J. and R. J. P. Schrama (1995). Identification and control - closed loop issues. *Automatica* **31**, 1751–1770.

Vandenberghe, L. and S Boyd (1996). Semidefinite programming. *SIAM Review* **38**, 49–95.

Walgama, K. S. (1986). Multivariable Adaptive Predictive Control for Stochastic Systems with Time Delay. PhD thesis. Dept. of Chem. Eng., University of Alberta. Edmonton.

Wise, B. M. (1991). Adapting Multivariate Analysis for Monitoring and Modeling of Dynamic Systems. PhD thesis. University of Washington.

Wold, P. R. (1966). *Multivariate Analysis*. Chap. Estimation of Principal Components and Related Models by Iterative Least Squares. Academic Press, NY.

Wolovich, W. A. and 1976 Falb, P. (1976). Invariants and cannonical forms under dynamic compensation. *SIAM J. Control* p. 996.

Wolovich, W. A. and H. Elliot (1983). Discrete models for linear multivariable systems. *Int. J. Control* **38**, 337.

Wood, R. K. and M. W. Berry (1973). Terminal composition control of a binary distillation column. *Chem. Eng. Sci.* **28**, 1707.

Wright, S. J. (1996). Applying new optimization algorithms to model predictive control. In: *Proceedings of the CPC-V*.

Wright, S. J. (1997). *Primal-Dual Interior-Point Methods*. SIAM. Philadelphia.

Zafiriou, E. (1990). Robust model predictive control of processes with hard constraints. *Comput. Chem. Eng.* **14**, 359–371.

Zafiriou, E. and A. Marchal (1991). Stability of siso quadratic dynamic matrix control with hard output constraints. *AIChE J.* **37**, 1550–1560.

Zang, Z. R., R. R. Bitmead and M. Gevers (1995). Iterative weighted least-squares identification and weighted lqg-control. *Automatica* **31**, 1577–1594.

Zheng, Z. Q. and M. Morari (1993). Robust stability of constrained model predictive control. In: *Proc. American Control Conf.*. San Fransisco, CA.

Zhu, X. and D. E. Seborg (1994). Nonlinear model predictive control based on hammerstein models. In: *Proceedings of PSE '94*.

Appendix A

A method for closed loop estimation of uncertainty

In chapter 5 we presented bounds on performance of an uncertain system. The knowledge of uncertainty was presumed. Here we present a non-parametric method to estimating the uncertainty in modelling under closed loop conditions, for open loop stable processes. The internal model control framework once again comes to the rescue, reducing the closed loop problem to the open loop one. Consider the IMC structure once again.We have the following relationships in the presence of feedback and assuming $r = v = 0$.

$$\begin{aligned} \hat{y} &= \hat{P}(I + Q\Delta)^{-1}w \\ d &= y - \hat{y} = (P - \hat{P})(I + Q\Delta)^{-1}w \end{aligned}$$

In terms of the relevant spectra we have:

$$\begin{aligned} \phi_{\hat{y}w} &= \left|\hat{P}(I + Q\Delta)^{-1}\right| \phi_w \\ \phi_{dw} &= \left|(P - \hat{P})(I + Q\Delta)^{-1}\right| \phi_w \end{aligned}$$

Dividing the second equation with the first we get

$$\frac{\phi_{dw}}{\phi_{\hat{y}w}} = \left|\frac{P - \hat{P}}{\hat{P}}\right| \triangleq |\Delta_M|$$

Consider the simulation example with the delay mismatch in chapter 5. We will apply the above method for estimating the frequency domain bound on the multiplicative mismatch. The true MPM in the multiplicative form is given by

$$\Delta_M = 1 - z^{-1}$$

for a delay mismatch of one sample. Figure A.2 shows the estimated mismatch along with the actual mismatch. The estimated mismatch shows a good fit except in the low frequency region.

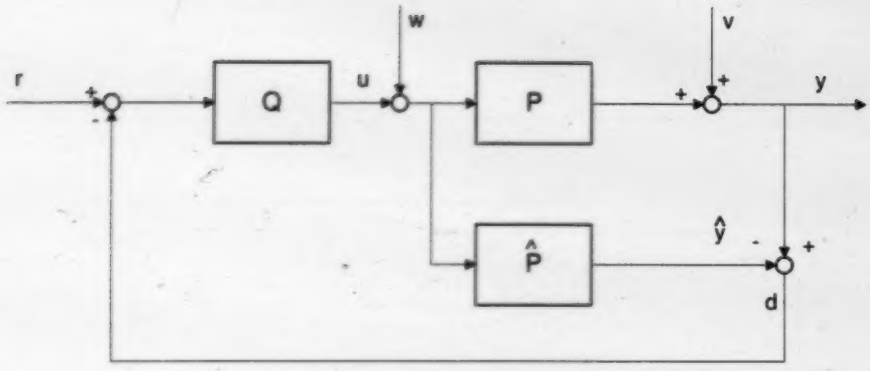


Figure A.1: *The IMC structure*

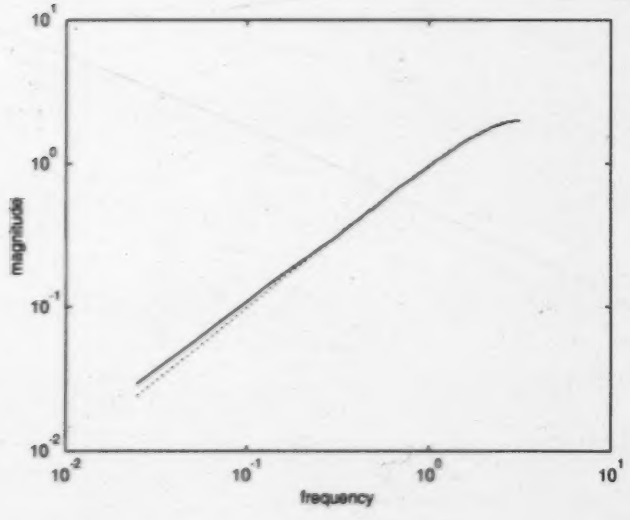


Figure A.2: *Comparison of the estimated mismatch (dashed line) with the actual mismatch (solid line).*